DNA Origami Nanostructures for the Delivery and Discovery of Nucleic Acids

Presented in fulfillment of the requirements for Undergraduate Research Distinction in Chemical and Biomolecular Engineering at The Ohio State University

Submitted by: Amjad Akif Chowdhury

Spring 2017

Faculty Advisor: Dr. Carlos Castro, Department of Mechanical and Aerospace Engineering

Co-advisor: Dr. Christopher Lucas, Post-Doctoral Researcher, Department of Mechanical and Aerospace Engineering

Thesis Committee: Dr. Carlos Castro & Dr. Jonathan Song, Departmental of Mechanical and Aerospace Engineering

Abstract

Scaffolded DNA origami is a novel DNA-based nanostructure fabrication method that boasts the advantages of ease of fabrication, high customizability in terms of geometry and modifications, and low cytotoxicity, making it an ideal candidate for biological applications. Several groups have already demonstrated its functionality as a drug delivery vehicle, delivering small molecule chemotherapeutics and proteins both *in vitro* and *in vivo*. However, the field is far from saturated, and many applications of these devices remain relatively unexplored. One such avenue within the scope of drug delivery is the delivery of antisense therapy, which is showing increasing promise in combating dysregulation of novel targets such as microRNA within disease states. Another approach is to expand the functionality of these nanostructures in biological systems to include the discovery of new interacting partners with ligands, such as microRNA, associated with the nanostructure. The purpose of this work is to expand the field of DNA origami into these fields. In the first project, nanostructures functionalized with sequences antisense to oncogenic microRNA on the outer surface were successfully fabricated. These structures were shown to successfully sequester their target microRNA selectively in solution, and could enter cells without any need of modification. Finally, preliminary data was collected suggesting that the amount to structures within cells can be quantified, and that these structures retain their functionality and are able to affect downstream processes *in vitro*. In the second project, a process schematic was developed to identify and extract ciRS-7, a circular RNA that strongly interacts with miRNA-7. Structures with overhangs containing the miDNA-7 sequence and a mechanism for binding to, and being released from streptavidin coated beads were successfully synthesized. Each step of the target discovery protocol was validated independently in solution.

This work will hopefully serve as both a basis for optimizing the methods and concepts included within, and as inspiration for the development of novel concepts that take the field even further beyond.

Acknowledgments

Firstly, I would like to thank Dr. Castro for the excellent mentorship and support that he has provided over the last three years. I was unsure as to what my career path should be when I joined his lab, and working for him has literally been a life changing experience.

I would also like to thank all the members of the Nanoengineering and Biodesign Laboratory (NBL), who have become very dear friends to me over the years. NBL provides a truly nurturing and amicable environment that I will be hard pressed to find anywhere else. They will truly be missed. In particular, I want to thank Dr. Christopher Lucas and Patrick Halley for being amazing mentors, who not only helped me with research but with planning for life as well.

I would like to thank all the members of the Byrd Lab, especially Dr. John Byrd, Dr. Amy Johnson, Dr. Ramiro Garson, and Dr. Yuh-Ying Yeh, for collaborating on these projects and providing crucial expertise without each none of this would be possible.

Special thanks to Laura Heyeck for editing this manuscript, and keeping me to task in general for this thesis.

Finally, I would like to thank my friends and family for supporting me through this entire process, and for lending a sympathetic ear whenever I needed to complain.

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. DNA Nanotechnology

Deoxyribonucleic acid, or DNA, is one of the fundamental building blocks of life. It is a polymer composed of four unique molecules known as nucleotides: adenine (A), thymine (T), cytosine (C), and guanine (G). Each nucleotide itself consists of three basic parts: a 5-carbon deoxyribose sugar ring, a phosphate group bonded to the 5' carbon of the sugar, and one of four nitrogenous bases attached to the 1' carbon of the sugar which gives the nucleotide its identity. The backbone of DNA is formed through condensation reactions between the 5' phosphate of one nucleotide and the 3' hydroxyl group of another. This property gives DNA strands directionality, from the free 5' phosphate to the free 3' hydroxyl. The nitrogenous bases can be classified into two groups based on chemical structure: the pyrimidines (T and C) which consist of a single ring, and the purines (A and G) which consist of two fused rings. The bases can form hydrogen bonding interactions with each other, and under standard conditions, the nucleotides between two strands of DNA will form Watson-Crick base pairs, i.e. A with T and C with G, with high specificity. This base pairing ensures that the greatest number of hydrogen bonds are formed while maintaining the isosteric geometry obtained by a purine-pyrimidine interaction. The strands are also aligned such that the base pairs are planar, thus maximizing π electron interactions between subsequent base pairs. This results in the double helical structure of DNA, shown in Figure 1.1, where the two strands run anti-parallel to each other, twisting around a central axis, with the bases facing towards the center of the helix and forming hydrogen bonds with its Watson-Crick complementary base. Within cells, the sequence of nucleotides encodes all the information necessary for cellular function, and the specific interactions between complementary bases allow DNA to be replicated with high fidelity for the transmission of genetic material to daughter cells. However, the same sequence diversity and interaction specificity can be co-opted for the assembly of highly programmable structures, giving rise to the field of DNA nanotechnology.

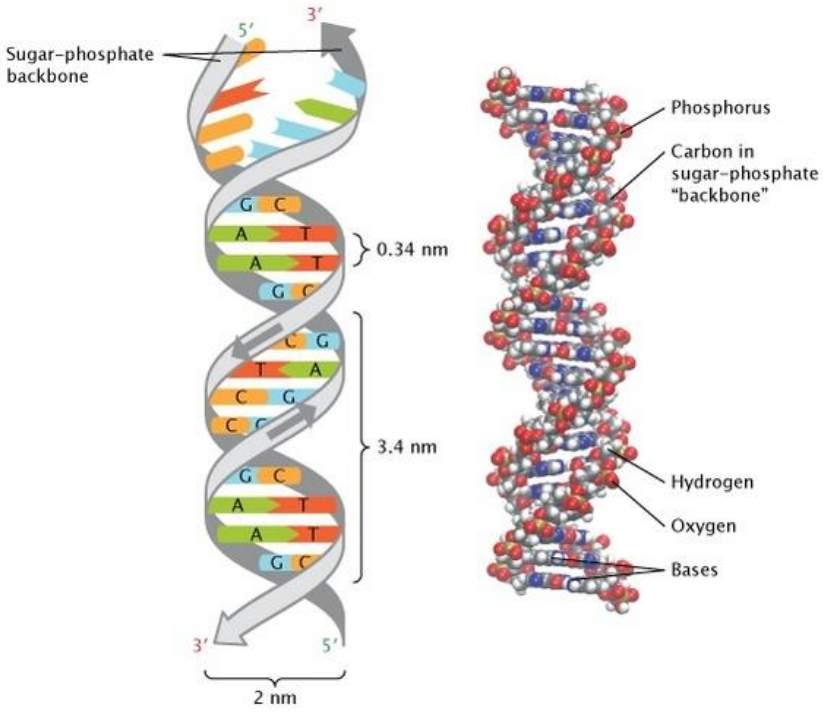

Figure 1. 1: Schematic representation (left) and atomistic model (right) of the DNA double helix[1].

The foundations of DNA nanotechnology were laid with Nadrian Seeman's work on characterizing immobile DNA junctions and constructing lattices of repeated junctions using sticky end adhesions[2]. This method was used to construct 3D structures such as cubes[3], as well as 2D tiled lattices and structures with different topologies[4]. However, since structure synthesis involved interactions between many short oligonucleotides, the stoichiometry of each strand had to be precisely controlled to maintain high yields, and multiple reaction and purification steps were necessary to form complex structures.

These limitations were addressed by Paul Rothemund, who devised an alternative fabrication method termed "scaffolded DNA origami", that consisted of a one-pot reaction between a long single stranded DNA (ssDNA) "scaffold" strand, and multiple shorter ssDNA "staple" strands[5]. Rothemund demonstrated the fabrication of 2D structures with highly programmable and precise geometries, and introduced patterns on the surface of his structures using DNA hairpins. Since then, the field of

scaffolded DNA origami has expanded rapidly to include 3D structures with various geometries and complex curvatures, and dynamic nanostructures demonstrating a complex range of motion (Figure 1.2).

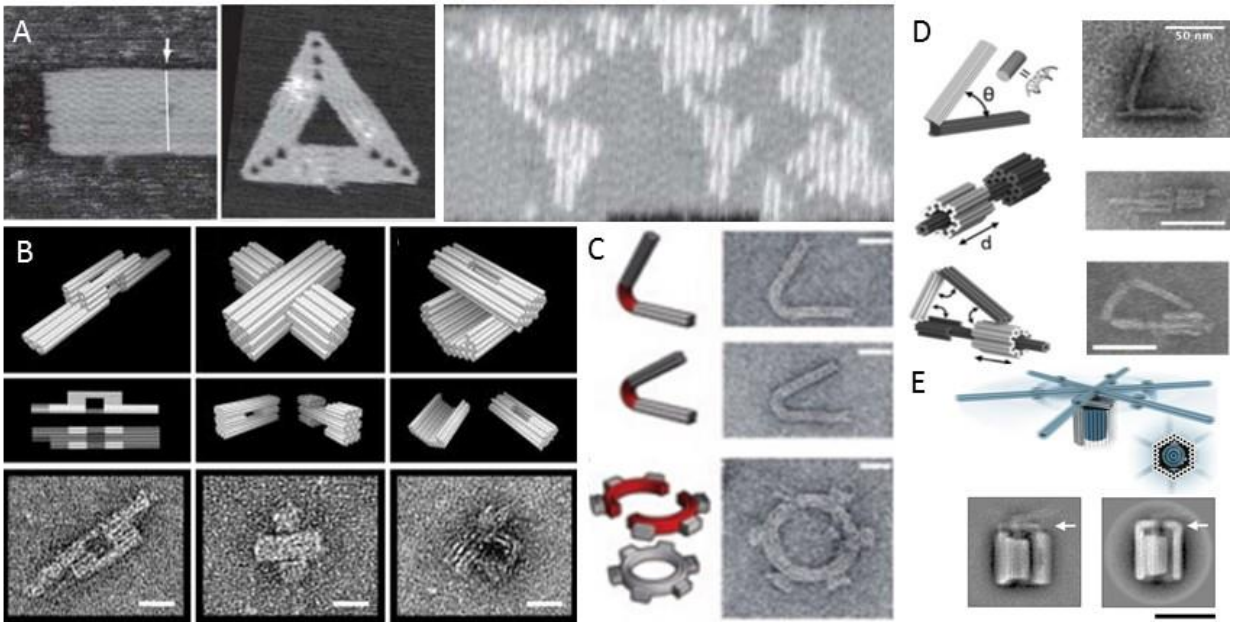

Figure 1. 2: Survey of DNA origami nanostructures. (A) 2D nanostructures with precise geometry and patterning[5]. (B) 3D origami nanostructures with different shapes[6] and (C) curvatures[7]. (D) 3D nanostructures with programmable angular and linear motions[8]. (E) Nanostructure with full rotational motion[9]. Scale bars: B,C: 20 nm, D,E: 50 nm.

While many applications of DNA origami nanostructures have been developed over the years, of particular interest is its function as a drug delivery vehicle *in vitro* and *in vivo*. DNA origami provides the advantages of high programmability and customizability, ease of fabrication, and inherent lack of toxicity, making it an ideal candidate for this role. For example, the clamshell structure in Figure 1.3(A) uses an aptamer based "lock and key" mechanism to identify target cells *in vitro* and selectively deliver the molecular payload contained within[10]. Nanostructures have also been used to deliver chemotherapeutics both *in vitro*[11] and *in vivo*[12], while circumventing drug resistance *in vitro*. Modifications have also been made to DNA origami nanostructures to improve their performance, such as in pharmacokinetics and biodistribution, in biological environments[13].

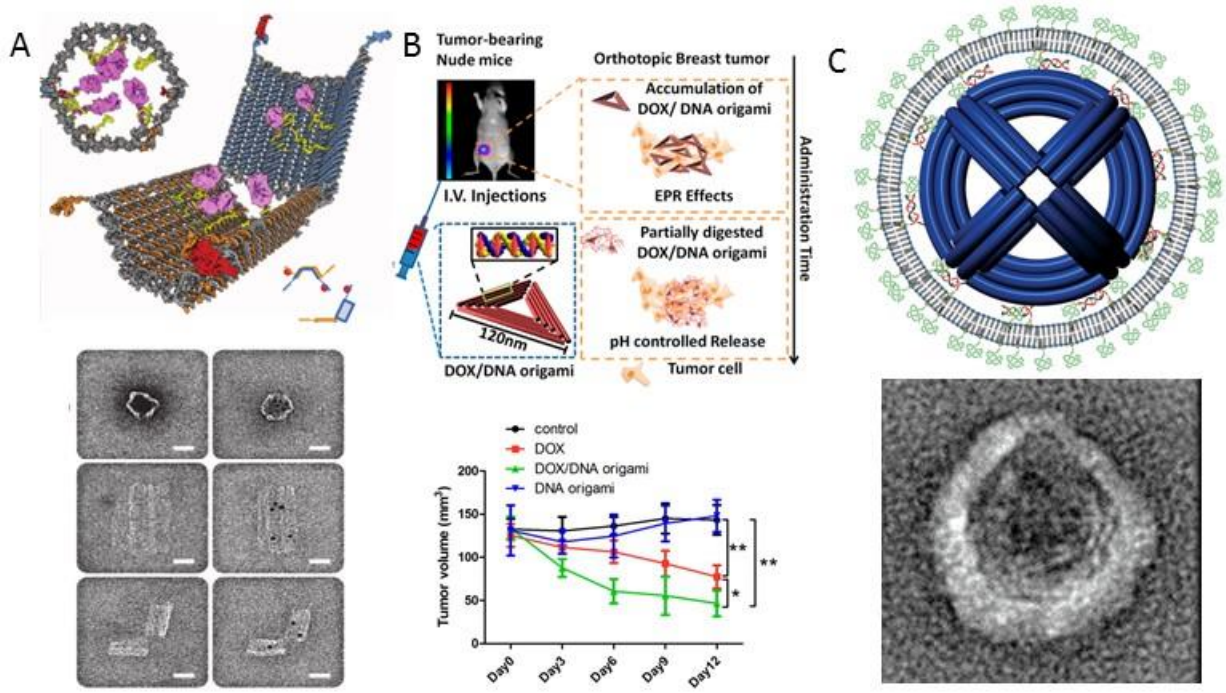

Figure 1. 3: DNA origami drug delivery vehicles. (A) Dynamic DNA nanostructure capable of delivering payloads in a targeted manner[10]. (B) Doxorubicin loaded DNA nanostructures showing efficacy in reducing breast cancer tumors in mice[12]. (C) Lipid coated nanoparticles for improved pharmacokinetics and biodistribution[13].

As mentioned earlier, the fabrication of DNA origami nanostructures involves the annealing of multiple short "staple" strands to one long "scaffold" strand. The staple strands are piecewise complementary to sections of the scaffold, and under annealing conditions the double stranded DNA (dsDNA) formed between the scaffold and staple gives it more structure. After this process is repeated with ~100 staples, the scaffold adopts its final desired structure. Computer aided design programs have been developed to aid in origami design[14], and certain rules have been developed to ensure proper incorporation of staples when designing these structures[15]. A schematic of this process is shown in Figure 1.4.

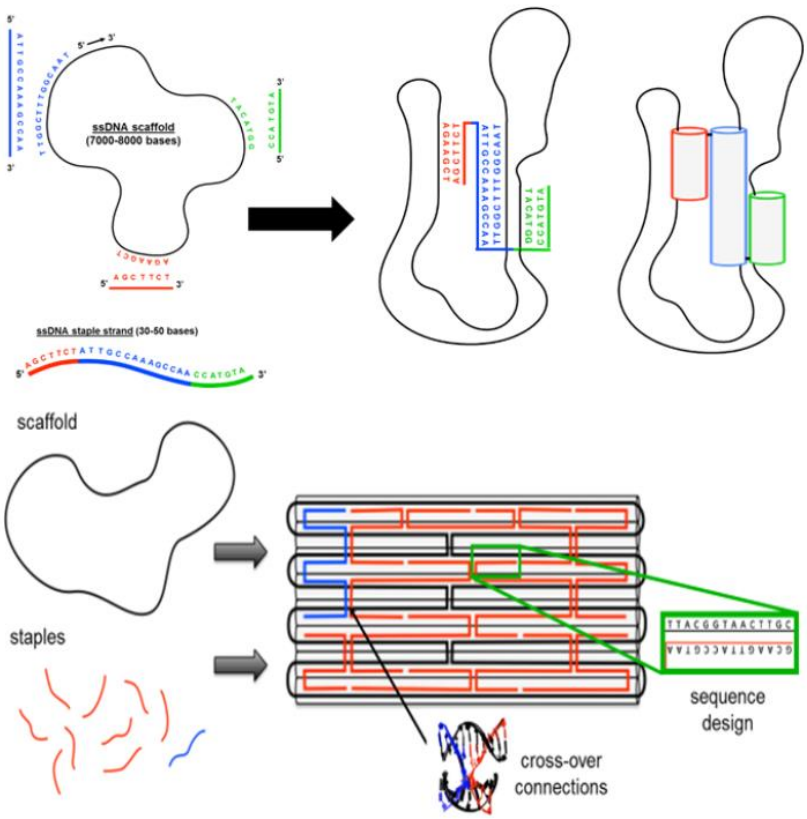

Figure 1. 4: Schematic of DNA origami folding. The top panel shows a piecewise complementary staple binding to the scaffold and bringing it together. dsDNA within the design are represented as cylinders for simplicity. The bottom panel shows a representative staple routing diagram highlighting the structure of staple crossovers.

The scaffold strand used is generally derived from the M13mp18 bacteriophage genome whose sequence is known, so the sequence of the staples can be generated using complementary base pairing rules. Traditionally, to fold nanostructures, the scaffold and staples would be combined in the proper ratio in addition to a buffer and divalent cations such as $MgCl_2$ to stabilize charged interactions, heated to 65°C, and then cooled slowly over a few days[15]. The initial heating ensures that existing secondary structure in the ssDNA have been broken before scaffold-staple interactions begin forming during the slow cooling. Recently, Sobczak et al have demonstrated that DNA origami structures can be folded at a constant annealing temperature over a period of hours rather than by slowly cooling it over a period of days, thus greatly simplifying the folding process[16].

1.2. *microRNA and circular RNA*

In addition to DNA, cells also contain a second form of nucleic acid, known as ribonucleic acid, or RNA. Chemically, RNA has two main differences with DNA: first, the 2' location on the ribose ring of RNA contains a hydroxyl group instead of a hydrogen group, and second, RNA molecules contain the base uracil (U) instead of T, which also forms a Watson-Crick base pair with A. However, the structure and function of RNA is very different from DNA. Firstly, RNA does not exist solely as a double helix, but can form a wide range of secondary and tertiary structures, similar to polypeptides. These conformations allow RNA to carry out functions other than information storage, such as target identification and reaction catalysis[17].

The most common forms of RNA are messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). mRNA is synthesized by transcribing, or copying the sequence of, DNA. It is then spliced to remove introns, or non-coding sequences, and is used to transfer information outside the nucleus. The mRNA then encounters the ribosome, which is composed of proteins and rRNA. The ribosome translates the message encoded in the mRNA into a sequence of amino acids, thus creating the polypeptide chain that is then folded into a functional protein. The amino acids are recognized and transferred to the ribosomes by tRNA[17]. However, in recent years, scientists have identified new types of RNA that were previously considered to be wasted byproducts of other reactions. One such RNA is microRNA (miRNA), short, single stranded, non-coding RNA molecules that were first characterized in 1993[18], and have since been implicated in a wide variety of diseases[19].

miRNA are initially transcribed and modified similar to mRNA, but contain multiple well defined secondary structures known as hairpins. These are identified and excised out of the longer strand, forming ~70 base hairpin structures that are exported out of the nucleus. In the cytoplasm, the hairpin is further processed until a ~22 base dsRNA strand remains. The dsRNA then dissociates and the mature single stranded miRNA associates with a protein complex known as the RNA induced silencing complex

(RISC). The miRNA is partially complementary to a region on its target mRNA known as the 3' untranslated region (3'UTR), and will bind to the 3'UTR if it comes into contact with the target mRNA. The RISC complex can then either act as a nuclease and degrade the mRNA, or can block ribosomes from translating the associated mRNA. In both cases, the translation of the mRNA into protein is suppressed[20]. Therefore, miRNA play a crucial role in the regulation of mRNA expression in cells. The biogenesis of miRNA is shown in Figure 1.5.

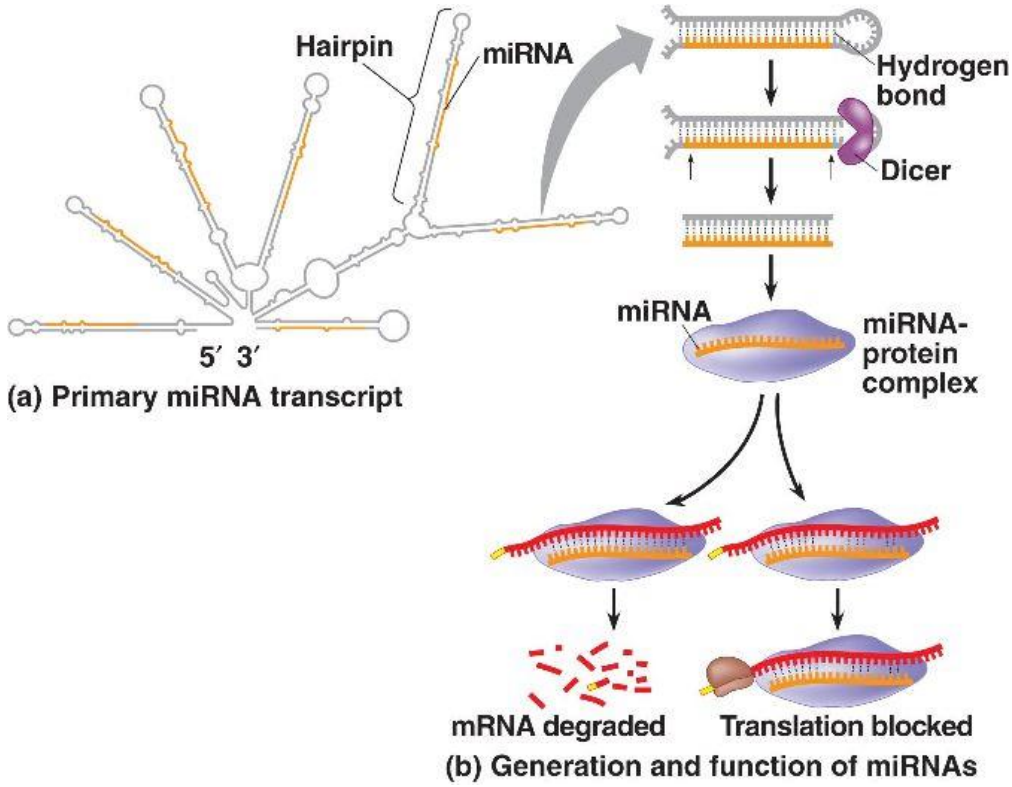

Figure 1. 5: Biogenesis of miRNA[17].

The dysregulation of miRNA have been implicated in many diseases, including cancer (Figure 1.6)[30]. In cancer, miRNA can serve as oncogenes by downregulating tumor suppressor genes, or as tumor suppressors by inhibiting oncogenes (Figure 1.6)[30]. Tumor suppressor miRNA are under expressed in cancerous tissue relative to healthy tissue, and oncogenic miRNA are overexpressed.

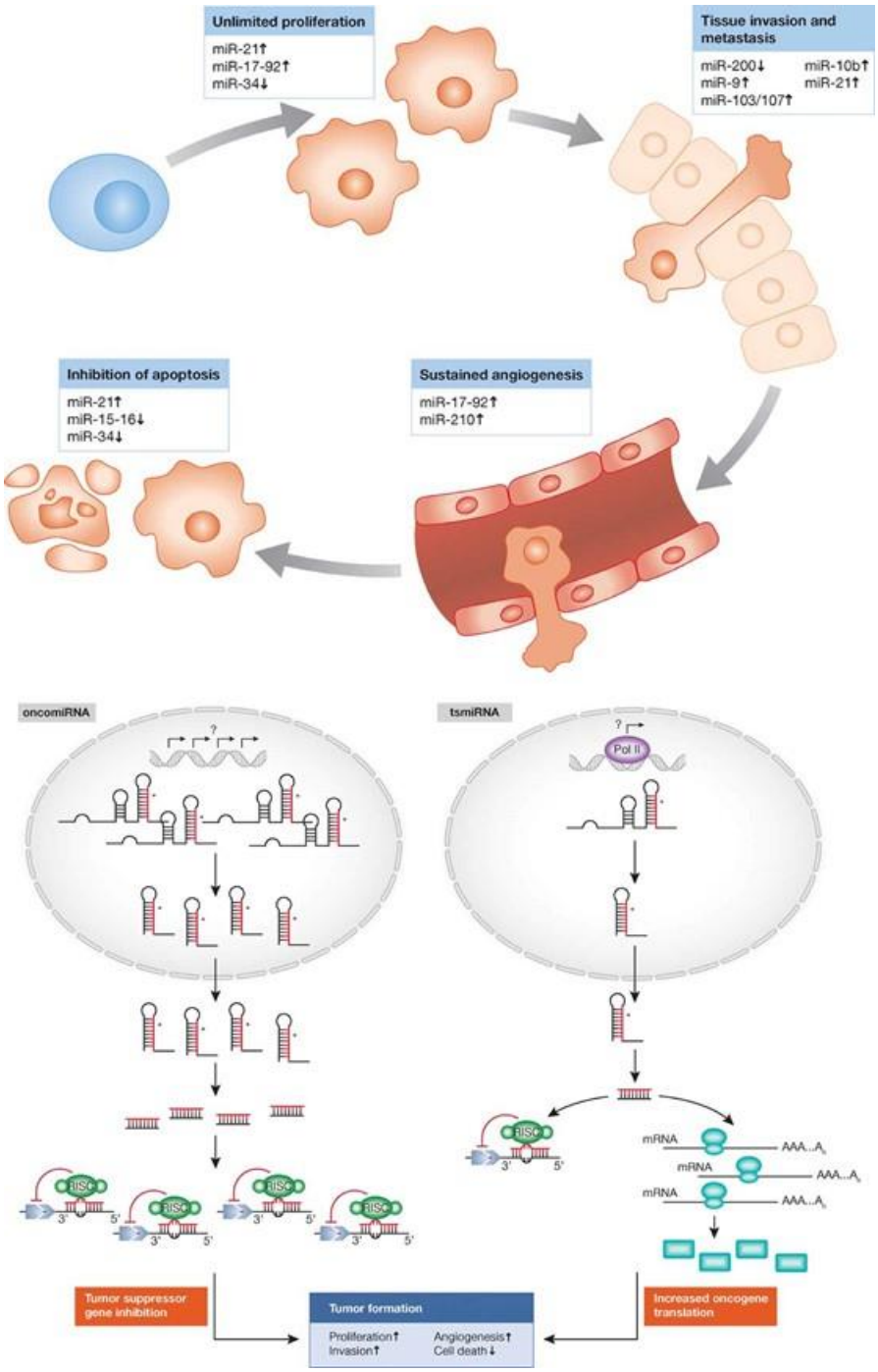

Figure 1. 6: (Top) miRNA associated with the different hallmarks of cancer. (Bottom) Mechanisms of action of both oncogenic miRNA and tumor suppressing miRNA[30].

For example, miRNA-155 (miR-155) has been shown to be overexpressed in Chronic Lymphocytic Leukemia (CLL)[21], and its overexpression has been linked to poor prognosis of patients[22]. While a single miRNA may regulate multiple pathways, one possible target for miR-155 is the protein

SHIP-1, which regulates the B cell receptor (BCR) pathway in B cell lymphocytes[23]. Another miRNA that is overexpressed in CLL[21] and also affects the BCR pathway[24] is miRNA-150 (miR-150). In addition to being overexpressed within the cellular environment, these miRNA are also secreted in high relative concentrations from the tumor cells into the surrounding environment within extracellular vesicles, or membrane bound vessels. These act as a means of communication between the cancer cells and its microenvironment, and may play a role in tumor immune response, metastasis, tumorigenesis, angiogenesis, drug resistance, or changes in signaling cascades[25].

As mentioned earlier, traditional mRNA production involves splicing the transcription product to remove introns. Splicing generally involves one end of an exon, or coding sequence, binding to a downstream end of the next exon while removing the intron in between, resulting in a linear product. However, in some cases the donor of an exon can bind to an acceptor upstream of itself, thereby creating a circular product, in a process known as back splicing (Figure 1.7)[26]. The resulting molecule is called a circular RNA (circRNA), and even though it was first identified in the 1990s, its function was then unknown.

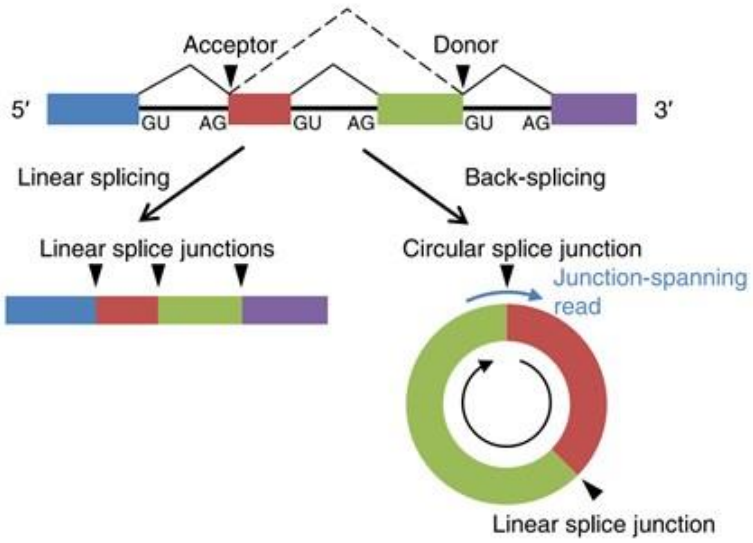

Figure 1. 7: Schematic showing linear splicing and back splicing to produce circular RNA[27].

In 2013, Memczak et al discovered one possible function of circRNA as miRNA sponges[28]. Specifically, they identified a particular circRNA they termed ciRS-7 that had multiple partially complementary binding sites for miR-7 that did not cause RISC formation and circRNA degradation. They therefore hypothesized that ciRS-7 can sequester miR-7 from solution, and prevent its regular function (Figure 1.8A,B). Alternatively, ciRS-7 also has binding sites for miR-671 that can cause RISC formation and ciRS-7 degradation, and so could trigger a release of sequestered miR-7, causing efficient repression of miR-7 targets (Figure 1.8C)[29]. However, no circRNA other than the ones reported by Memczak et al have shown anti-miRNA activity, so the function of circRNA is still an open question[27].

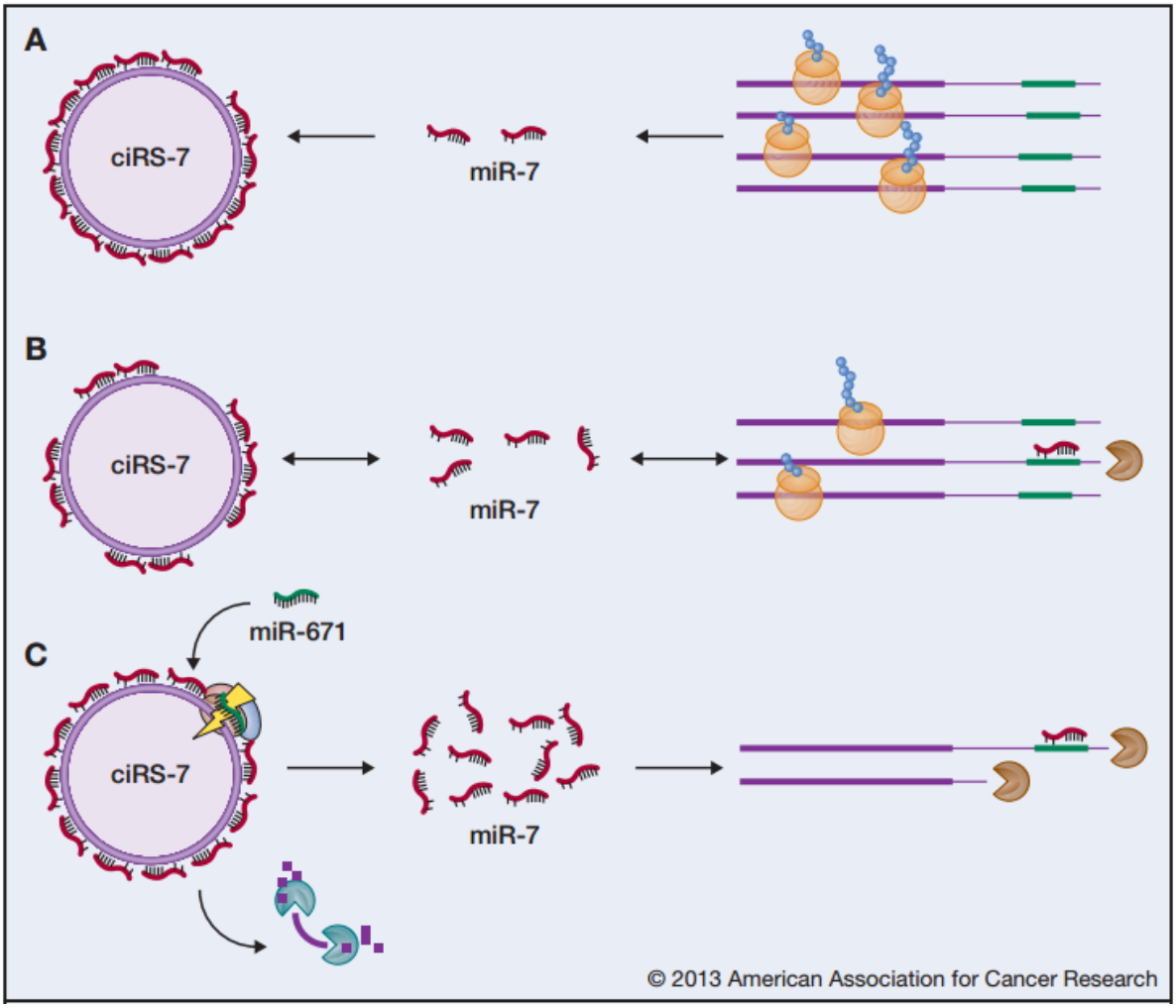

Figure 1. 8: Possible mechanisms of ciRS-7 function. (A) ciRS-7 can act as a miR-7 sponge, and prevent miR-7 activity. (B) ciRS-7 can act as a miR-7 buffer and modulate miR-7 activity. (C) ciRS-7 can act as a miR-7 reservoir and promote miR-7 activity[29].

1.3. Objectives

The objectives of this thesis are twofold. The first objective is to expand the repertoire of drug delivery candidates for DNA origami to include anti-miRNA therapy due to its potential in treating many different forms of cancer. In collaboration with Dr. John C. Byrd of The Ohio State University Comprehensive Cancer Center, we aim to develop a nanostructure capable of sequestering the overexpressed miR-155 in an *in vitro* CLL model as a proof of principle.

The second objective is to develop a methodology of the detection and extraction of macromolecules associated with miRNA from cells to identify novel interacting partners of miRNA. As a proof of principle, we aim to demonstrate the successful extraction of ciRS-7 using nanostructures expressing miR-7.

2. Structure Design and Characterization

2.1.Structure Design

The nanostructure used in this thesis was adapted from an earlier design by Halley and Lucas et al[32] and was termed the "Branch". The Branch is a rod-shaped 26 helix bundle structure roughly 100 nm in length with internal voids, similar to the structure by Halley and Lucas et al (Figure 2.1). The structure uses the 7249 base M13mp18 genome as its scaffold. The staple routing and staple sequences were generated using the software cadnano[14] and is included as Appendix A. All staples were ordered from Eurofins MWG Operon USA (Louisville, KY).

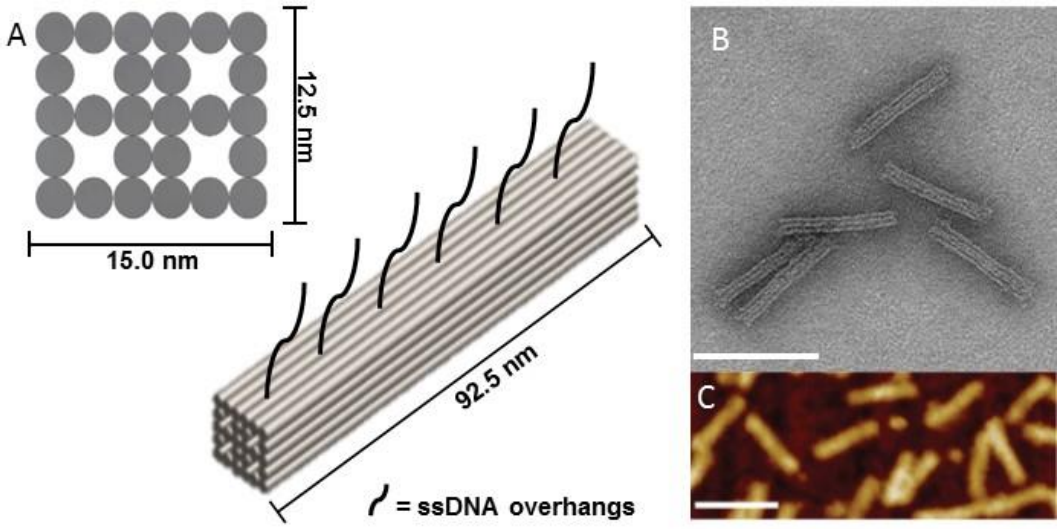

Figure 2. 1: (A) Solid model of the Branch showing dimensions and overhangs. (B) Transmission electron micrograph (TEM) of the Branch. (C) Atomic force micrograph of the Branch. Scale bars: 100 nm.

The key difference between the two designs is the presence of 42 ssDNA "overhangs" patterning the outer surface of the Branch. These overhangs were created by extending the staples on the outer helices beyond their complementary scaffold sections such that they were exposed to the solutions. The arrangement of these overhangs on each face of the Branch is shown in Figure 2.2. The overhangs locations were chosen such that they were spaced evenly while ensuring that a 3' staple end was available at that position.

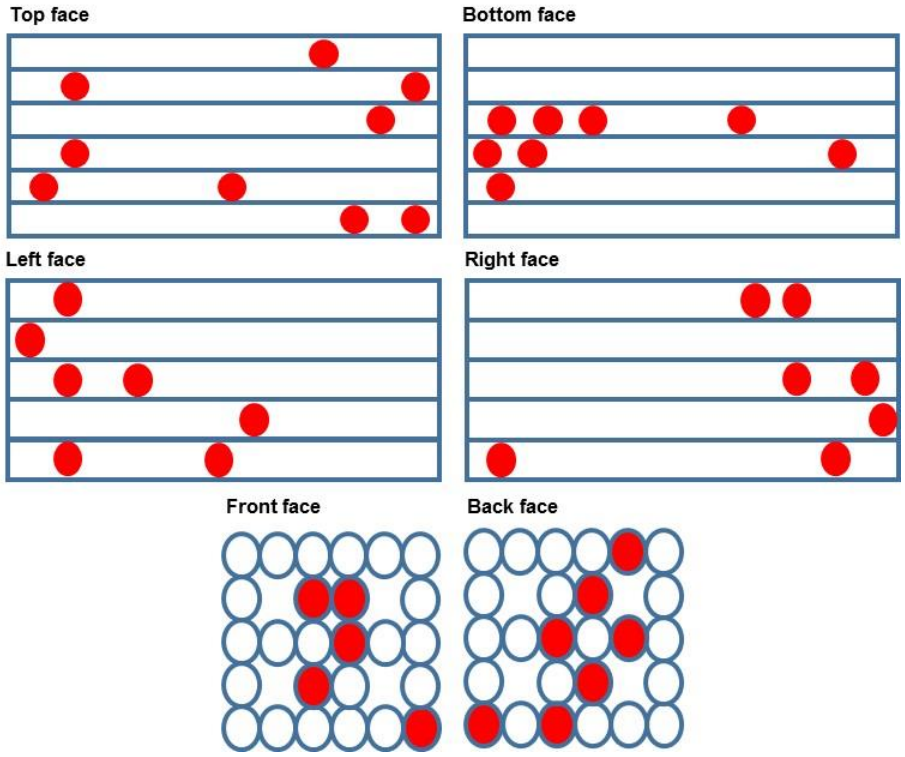

Figure 2. 2: Overhang positions on the Branch.

The overhang sequences are completely customizable and they serve as the functional component of the structure. The different versions of the Branch fabricated and their overhang sequence are listed in Table 2.1.

Structure	**Function**	**Overhang sequence (5'->3')**
No overhang (noOH)	Control	None
Scrambled (scr)	Control	Variable
Anti-miR-155 (am155)	Antisense	TTTTTACCCCTATCACGATTAGCATTAA
Anti-miR-150 (am150)	Antisense	TTTTTCACTGGTACAAGGGTTGGGAGA
miDNA-7 (D7)	DNA presentation	TTTTTTGGAAGACTAGTGATTTTGTTGT
	Biotin attachment	TTTTTTAACCGCTCTGGTTAACGTGTCTGGGCAAAA

Table 2. 1: List of different structures by overhang sequence.

The noOH structure does not contain any overhangs, and is used to control for the effect of having overhangs on the structure. The scr structure contains 42 overhangs, each with a unique, random sequence, and is used to control for non-specific effects of the overhangs. The sequence of each scrambled overhang is included in Appendix A. The final three are functional nanostructures, with the sequence marked in red denoting the functional sequence. Both am155 and am150 contain DNA sequences that are fully complementary to their respective target miRNA, and are used in the anti-sense project. D7 contains two different types of overhangs. 41 out of 42 overhangs contain the DNA analogue of the miR-7 sequence for ciRS-7 binding, shown in red in Table 2.1. The red section of the final sequence is complementary to an ssDNA strand containing a biotin molecule, and the blue portion of the sequence denotes the toehold region for strand displacement. This structure is used in the target discovery project and will be elaborated on in Chapter 4.

2.2.Structure Characterization

Multiple folding parameters of the structures were characterized to optimize the folding process. Initial characterization was carried out using the noOH structure. The first characterization was done by imaging structures using transmission electron microscopy (TEM) and atomic force microscopy (AFM). For TEM, structures were deposited on copper mesh grids and stained with uranyl formate for imaging. For AFM, structures were adsorbed onto mica surfaces for imaging. As shown in Figure 2.1B-C, the structures look well folded and adopt the expected geometry.

Next, the optimum salt $MgCl_2$ concentration for structure fabrication was determined. Structures were folded at $MgCl_2$ concentrations ranging from 12 mM to 26 mM in a slow cooling ramp[15] and then run on a 2% agarose gel in the presence of 11 mM $MgCl_2$ at 70 V for 2 hrs. The resulting bands were compared to the scaffold band and imaged using TEM (Figure 2.3A). The structures folded successfully over a long range of $MgCl_2$ concentrations (12 – 20 mM) without any aggregation, with optimum folding at 18 mM concentration.

Next, the ideal annealing temperature for structure folding at a single temperature[16] was determined. The mixture of scaffold and staples with 18 mM $MgCl_2$ was heated at 65.0°C for 15 min, and then held at a constant temperature between 60.0°C and 40.0°C for 4 hrs, before being cooled down to 4.0°C. The structures were then run on an agarose gel (Figure 2.3B). The Branch folded over a wide range of annealing temperatures starting at 58.8°C. To obtain a finer estimate of the upper limit of annealing, the same experiment was conducted with a temperature range of 62.0°C and 58.0°C, which showed that the Branch was successfully folded at 59.6°C (Figure 2.3C). Since the Branch seemed to fold at any temperature between 59°C and 40°C, a temperature of 54°C was chosen for subsequent folding reactions.

Finally, the folding kinetics of the Branch were determined. After the 65°C melting step, the folding reaction at 18 mM $MgCl_2$ was held at 54°C for varying amounts of time between 1 and 120 min, before being quenched on ice. The samples were then run on an agarose gel. The nanostructure began to fold as early as 1 min into the reaction, with complete folding achieved by 5 min (Figure 2.3D). An annealing time of 1 hr was chosen for subsequent folding reactions.

There were some discrepancies observed between the noOH structure and the structure containing overhangs. First, the addition of overhangs of any sequence retarded the bands on the gel. This is possibly due to the added weight and resistance imparted by the unstructured ssDNA when flowing through the gel. However, none of the folding characteristics studied were impacted by overhang incorporation, and structure yields were not adversely affected.

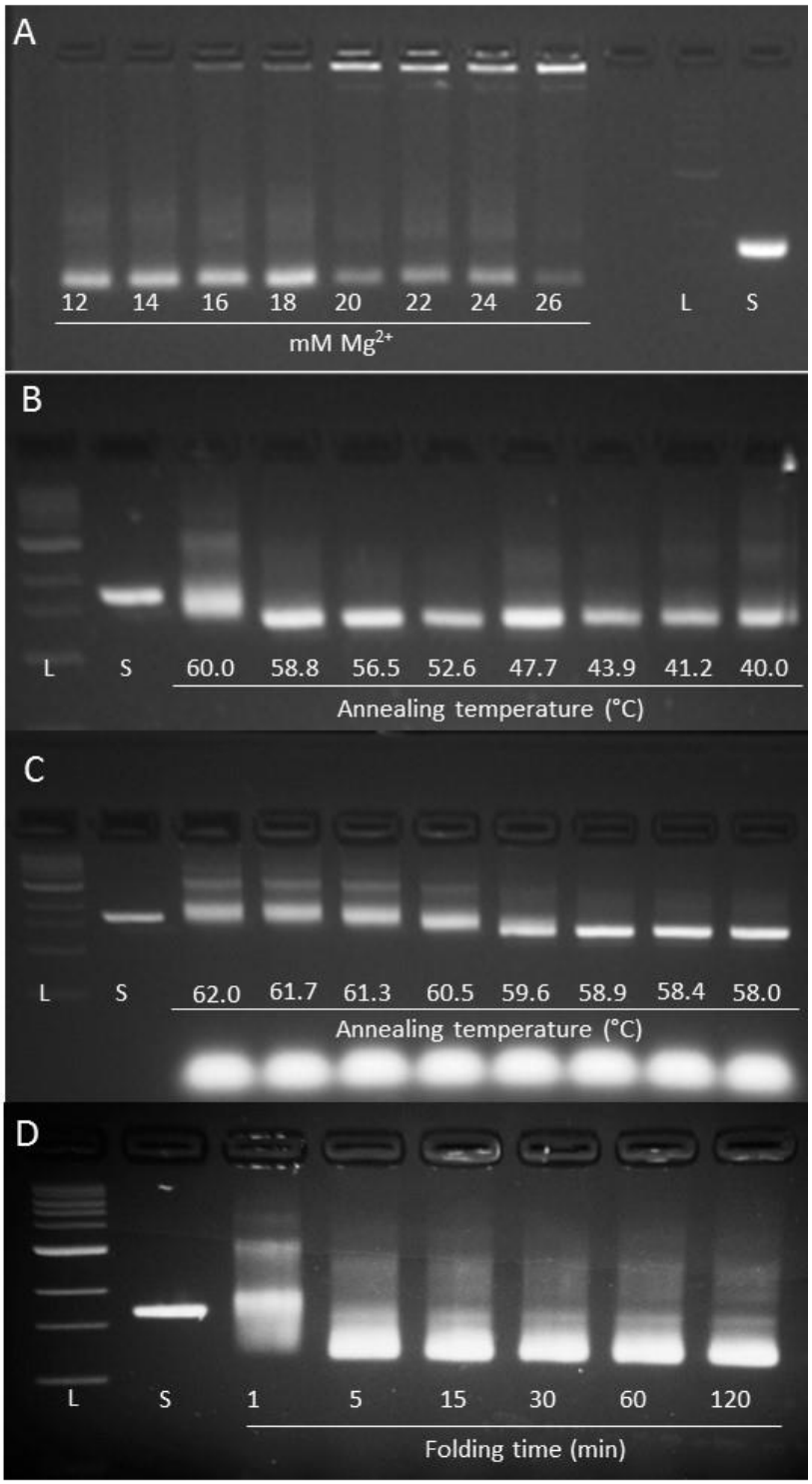

Figure 2. 3: Folding characterization of the Branch: (A) salt screen, (B) coarse grain annealing temperature (C) fine grain annealing temperature (D) folding time. L: ladder; S: scaffold.

3. Antisense Delivery

3.1. Overview

As stated earlier, the objective of this project was to design and validate a nanostructure capable of sequestering miR-150 and miR-155 within CLL to reverse their oncogenic effects. The am155 and am150 structures were designed for this purpose. The validation of these structures was broken down into three parts: selective sequestration of target miRNA in solution, confirmation and quantification of cellular uptake, and sequestration of target miRNA *in vitro*.

3.2. Sequestration in Solution

The selectivity of miRNA sequestration in solution was verified using a fluorescent sequestration assay, schematically shown in Figure 3.1. miR-155 conjugated with the ATTO550 fluorophore (fmiR-155), and miR-150 conjugated with ATTO647 fluorophore (fmiR-150) (IDT, Coralville, IA) were mixed in different proportions in solution. Then, either am155 or am150 structures were incubated with the mixture at a final concentration of 1 nM. The structures were incubated in the mixture at 37°C for 2 hrs. Afterwards, they were run on an agarose gel and fluorescently imaged using a Typhoon FLA 9500 (GE Healthcare and Life Sciences, Chicago, IL). The gel was then stained with SYBR Gold total DNA stain (Thermo Fisher Scientific, Waltham, MA) and imaged.

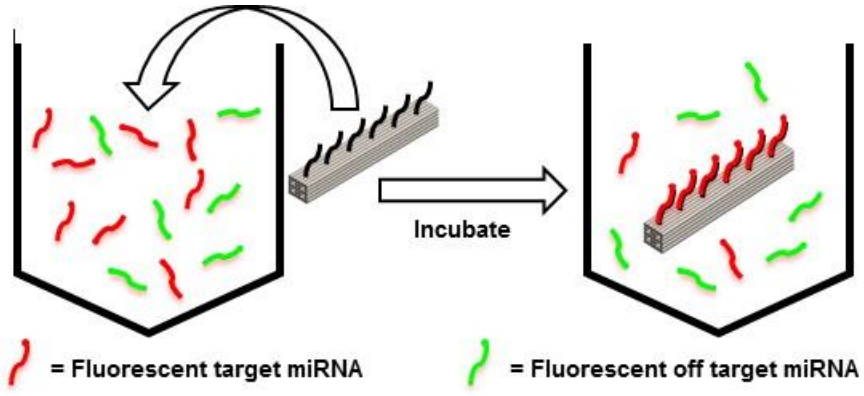

Figure 3. 1: Schematic of miRNA sequestration assay.

Since the overhangs on the structures are fully complementary to their target miRNA, we hypothesized that only the target miRNA would bind to the structure. Therefore, the structure bands would appear on the gel only when the gel is being imaged by the laser that excites the target miRNA fluorophore. For example, when the am155 structure is placed in a mixture of fmIR-155 and fmiR150, only fmiR-155 should bind to the structure, and the structure bands should only be visible when the gel is being imaged using the 532 nm laser. We further hypothesized that, as the relative concentration of target miRNA increases, the signal intensity from the structure bands should also increase, until it is saturated by miRNA.

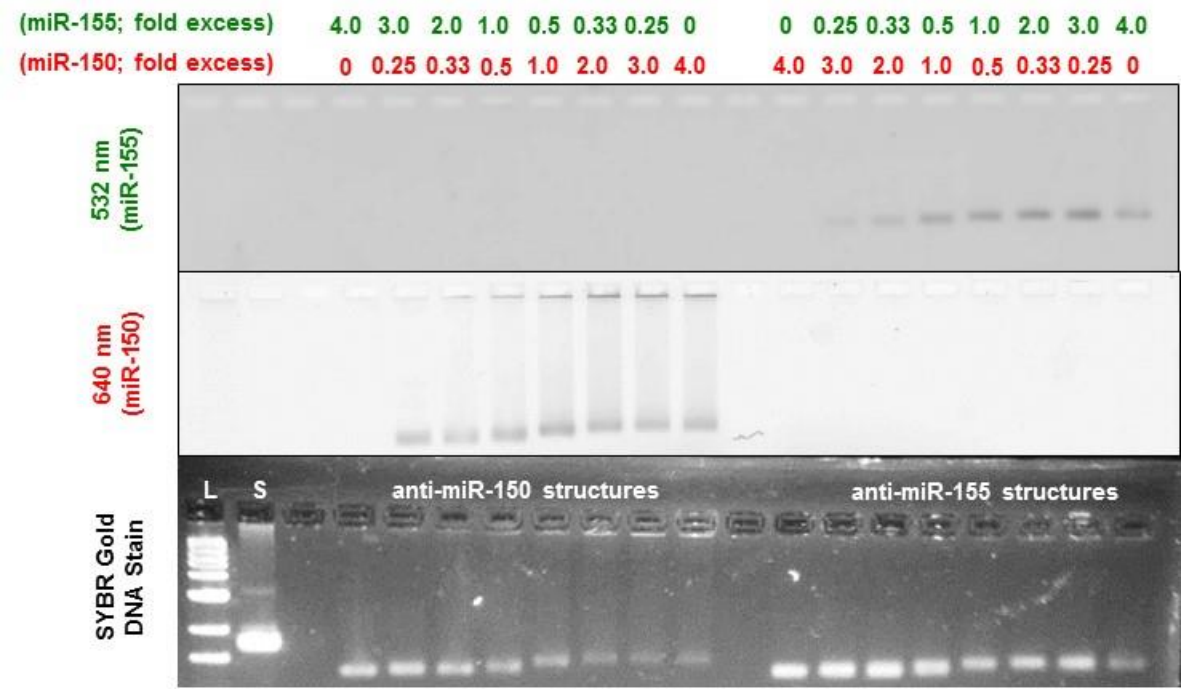

Figure 3. 2: miRNA sequestration assay. Each sample well contains both fmiR-155 and fmiR-150 in different ratios of fold excess relative to the number of overhangs on the structure, shown at the top. The sample block on the left all contain 1 nM am150 structures, and the sample block on the right all contain 1 nM am155 structures. The top gel image was taken using the 532 nm laser, which excites fmiR-155. The middle gel image was taken using the 640 nm laser, which excites fmiR-150. The bottom gel image was taken on a 480 nM light table, and shows the SYBR Gold total DNA stain.

Figure 3.2 shows representative results of the miRNA sequestration assay. As expected, when the gel was imaged using the green laser, the only signal came from the am155 structures, and when the

gel was imaged using the red laser, the only signal came from the am150 structures, suggesting that the structures were only able to sequester their respective target miRNA.

Interestingly, there is an evident gel shift of the bands as the concentration of the target miRNA in the sample increases until a 2 fold excess, after which the gel shift stops. This is possibly due to the increased mass of the structure band from the miRNA, which slows the band down. After 2 fold excess, the structures are saturated with miRNA, and so adding further miRNA in solution does not affect the mass of the structure, and so the gel shift stops. This was not expected, and suggests that the effect of sequestration can be observed without necessitating any fluorescence measurements.

Regardless, the fluorescent intensities of the structure band at different target miRNA concentrations were obtained by measuring the pixel intensities of each well in the gel image using a MATLAB script (included in Appendix B) and then normalizing by total mass of DNA in the well using the SYBR Gold total stain. The intensity values were then plotted against target miRNA fold excess, and are shown in Figure 3.3.

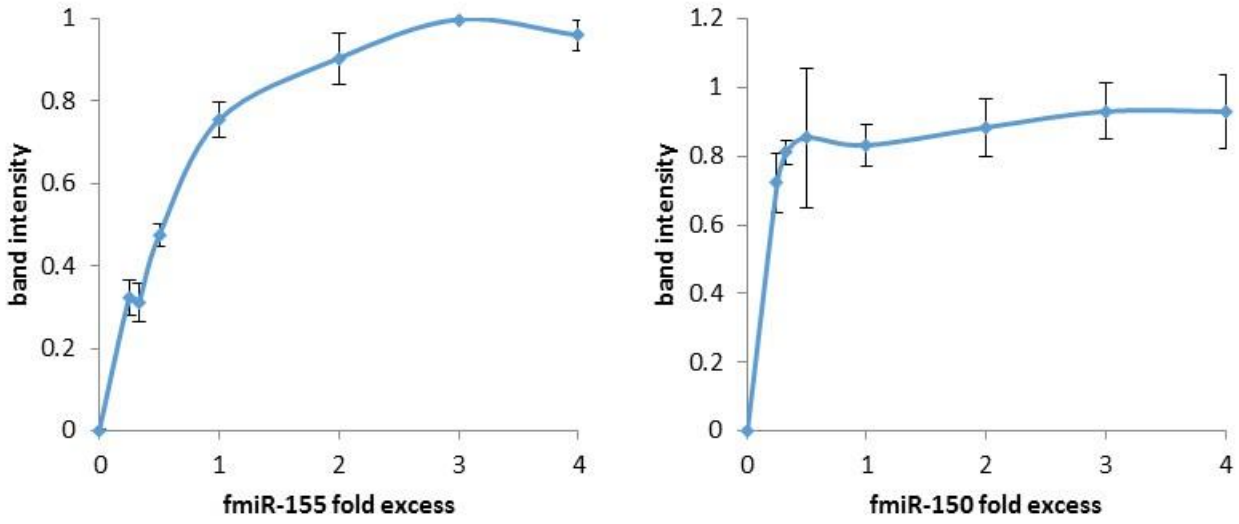

Figure 3. 3: Densitometric analysis of gel band intensities for am155 with fmiR-155 (left) and am150 structures with fmiR-150 (right). n=3

The gel intensity curves follow the same overall pattern as the gel shifts, with intensity increasing as fold excess target miRNA increases, until the structures are saturated at around 2 fold excess. The am150 curve (Figure 3.3, right) is less smooth compared to both the am155 curve (Figure 3.3, right) and the gel shift progression. This may be due to errors in the normalization to the SYBR Gold total DNA image. Since the concentration of structures was so low, the gels had to be stained in SYBR Gold solution for over 45 min, causing the signal to noise ratio to decrease and random bright spots to appear on the gel. This made it difficult to accurately quantify the total DNA present in the sample, which may have affected the intensity calculations.

To ensure that the fluorescent signal in the well was coming from the fluorescent miRNA bound to the structure and not from another source, a signal specificity assay was performed (Figure 3.4). Solutions were prepared containing a mixture of the same fluorescent and non-fluorescent miRNA (cmiR-155 or cmiR-150) (IDT, Coralville, IA), at different ratios. 1 nM of structures were incubated in the solution containing its target miRNA for 2 hrs at 37°C. After incubation, the structures were run on a gel, and the gel was imaged fluorescently and after staining with SYBR Gold. If the signal is specific to the fluorescent miRNA, we hypothesized that as the concentration of non-fluorescent miRNA increases, the intensity of the signal from the well should decrease.

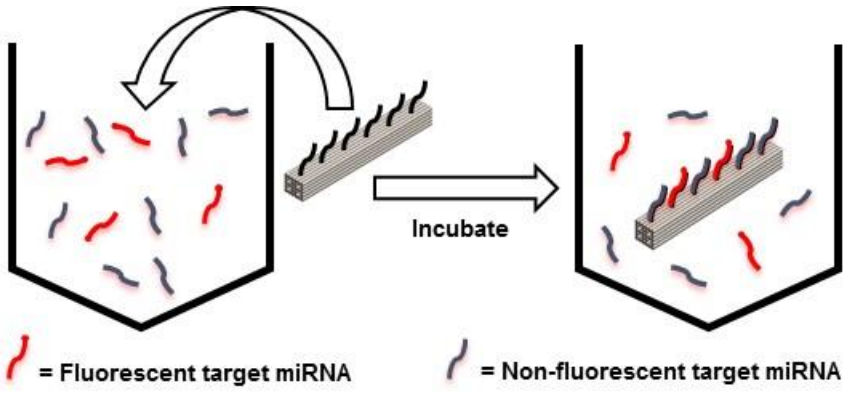

Figure 3. 4: Schematic showing methodology for signal specificity assay.

Figure 3.5 shows representative results from a signal specificity assay. As expected, the intensity from the bands decreases as the fold excess of non-fluorescent miRNA increases for both structures. There is no shift in the bands on this gel, which is expected since any spot not occupied by a fluorescent miRNA is expected to be occupied by a non-fluorescent miRNA, so the mass of structures in every well should be the same. The densitometric analysis shows that even though the signal was effectively gone at 10 fold excess cmiR-150 for am150 structures, there was still a noticeable signal in the am155 sample at the same concentration. This could suggest a difference in binding strengths between the fmiR-155 and cmiR-155. However, more experiments are needed to draw any conclusions. The assay should also be performed with 20 fold excess cmiR-155 for a more direct comparison between the two.

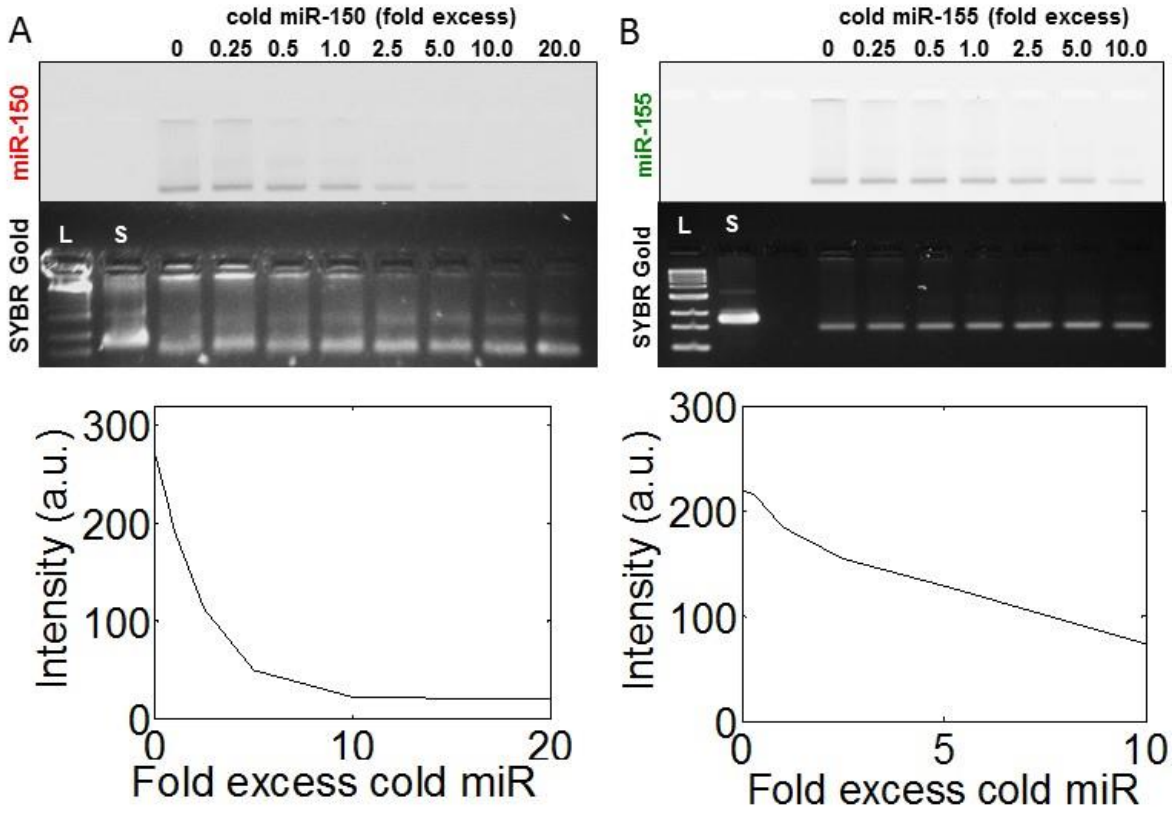

Figure 3. 5: Signal specificity assay for (A) am150 structures with fmiR-150 and cmIR-150 and (B) am155 structures with fmiR-155 and cmiR-155. The top gel shows the fluorescent image taken with (A) 647 nm laser and (B) 532 nm laser. The bottom gel shows the SYBR Gold total DNA stain. The graphs show densitometric analyses of the gel band intensities for the two gels.

3.3. Cellular Uptake

Cellular uptake was first assessed qualitatively using fluorescence microscopy experiments using the OSU CLL cell line obtained from the Byrd Lab (Figure 3.6). The internal vesicles within these cells were stained with a Lysotracker Green dye (Thermo Fisher Scientific, Waltham, MA), which fluoresces when excited with 480 nm light. The noOH structures were used for these experiments and were stained with TO-PRO-3 dye (Thermo Fisher Scientific, Waltham, MA), which intercalates into DNA and fluoresces when excited by a 647 nm laser. The labeled structures were then incubated with the labeled cells for varying periods of time, and imaged using a fluorescent microscope (Nikon, Tokyo, Japan).

It is possible, nominally, to colocalize brightfield images of cells with the fluorescent signal from structures to determine whether the structures are within the cells or not. However, since our image is only a 2D slice, it is difficult to determine whether the fluorescent signal from the structure arises from within the cell or from the surface of the cell. Furthermore, it is believed that DNA origami structures enter the cells through an endolysosomal pathway, in which the structure binds to receptors on the cell which causes the cell membrane to encapsulate the structure and form an internal vesicle[32]. Therefore, if there is colocalization of signal between the nanostructures and the endosomes, then we can conclude that the signal is coming from within the cells. Our hypothesis was that we will observe colocalization of signals between the Lysotracker Green and TO-PRO-3, suggesting that structures have entered the cells, possibly through an endolysosomal pathway.

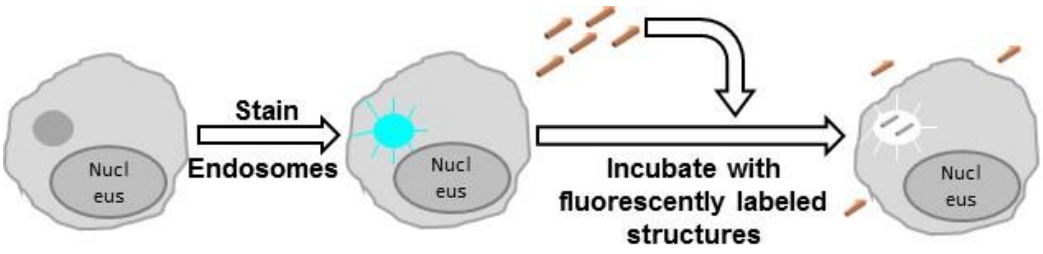

Figure 3. 6: Schematic of qualitative cellular uptake experiment using a fluorescence microscopy assay.

Experiments were first conducted with 0.5 nM final concentration of nanostructures in each well, incubated for 4 hrs. However, very little colocalization was observed, suggesting that either the structures were not entering the cells, or there were not enough structures present for a robust response. Therefore, the concentration of nanostructures was increased to 10 nM final concentration, and the incubation time was increased to 8 hrs. Representative results are shown in Figure 3.7. The figure shows colocalization between the Lysotracker signal and the TO-PRO-3 signal, suggesting that structures were taken up by the cells.

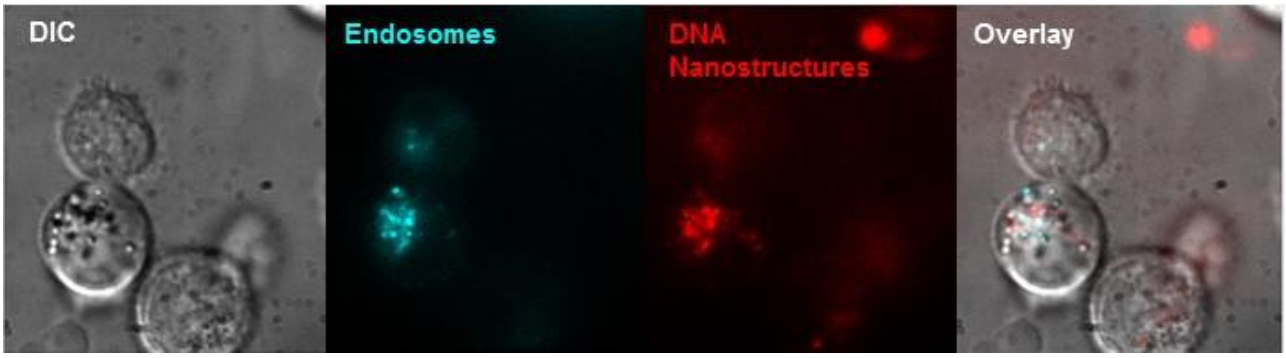

Figure 3. 7: Fluorescence microscopy images of cellular uptake of DNA origami. ~100,000 cells labeled with Lysotracker were seeded in an 8 well imaging plate, and structures labeled with TO-PRO-3 at 10 nM final concentration were incubated with cells for 8 hrs. The samples were then imaged with the fluorescent microscope. The white in the far right represents colocalization of signal.

While this method provides strong visual evidence for cellular uptake, it does not provide an easy way to quantify the amount or efficiency of uptake. Therefore, a quantitative polymerase chain reaction (qPCR) assay was developed to better quantify our results.

PCR (Figure 3.8) is a method by which a target DNA sequence can be amplified exponentially and specifically. The basic reaction mix for PCR contains the template, single nucleotides, the polymerase enzyme which synthesizes DNA, and primers. The template consists of the target DNA to be amplified and possibly off-target DNA as well. Primers are ~20 base length ssDNA that are complementary to the ends of the target DNA sequence. The reaction mixture is heated to 95°C to remove any secondary structure interactions, and then cooled to an annealing temperature, where the primers can specifically

bind to the template DNA. It is then heated up to a temperature where the polymerase enzyme has optimal activity. The polymerase then binds to the primer-template complex and extends the primer to create the complementary strand to the template. The denaturation-annealing-extension process is repeated tens of times. At 100 efficiency, the number of only the dsDNA template will double, thereby amplifying the template at an exponential rate until the starting material is used up and the reaction saturates.

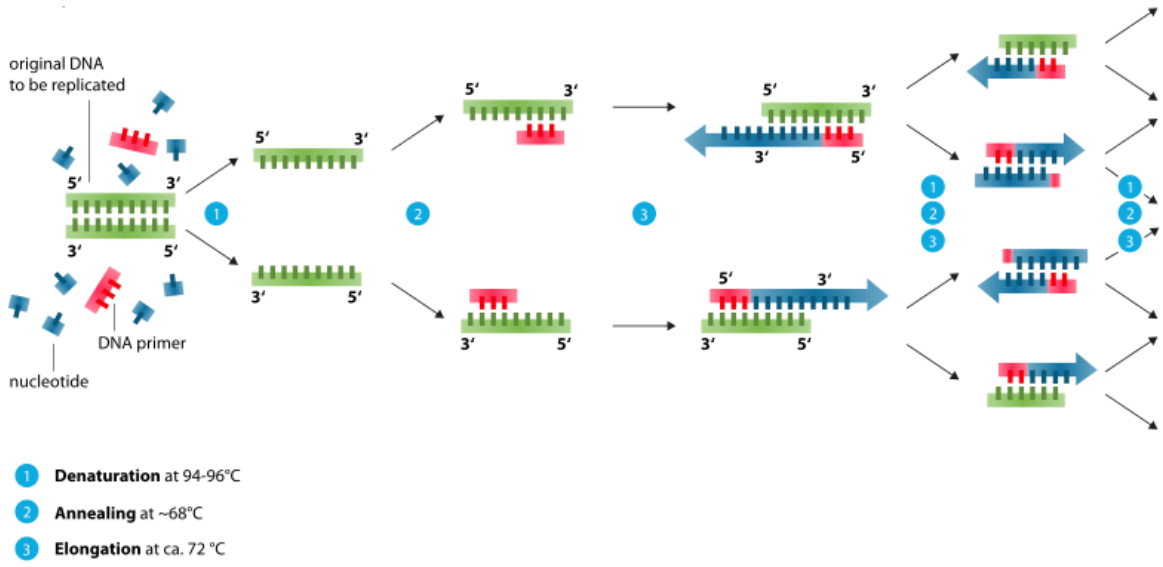

Figure 3. 8: Schematic for PCR.

For quantitative PCR, the reaction is the same, but the mixture also contains at least one fluorescent dye as well. In this version, the thermocycler in which the qPCR is being performed can excite the sample and take fluorescent measurements, and it does so after every cycle. The dye used emits light at different intensities based on whether it is interacting with ssDNA or dsDNA, and so as the concentration of dsDNA increases, the fluorescence intensity from the sample increase proportionally. Therefore, the method can be used to identify the relative concentration of dsDNA in solution at any given cycle number. Since the rate of production of target DNA is proportional to the starting concentration of the target DNA, the cycle number at which the target DNA achieves an arbitrary

fluorescence value can be used to determine what the starting concentration of target DNA was. To quantify the starting amount in absolute terms, a standard curve must first be created by amplifying solutions containing known quantities of target DNA and recording the cycle number at which the solution reached the set threshold. This curve can then be used to quantify the concentration of an unknown sample. Once the amplification is complete, the sample can be slowly heated to 95°C, and the fluorescence signal can be monitored over time. If there is only DNA of one length, as is expected from a correct amplification, then all the DNA strands should melt at the same temperature, resulting in a significant drop in fluorescence intensity at that temperature. If there are multiple intensity drops from a sample, then it can be concluded that the amplification was not specific.

For our system, the goal was to extract the total DNA from cells incubated with our structure and amplify the samples using qPCR to detect and quantify any DNA from our nanostructure, such as the scaffold, that may have entered the cell. We hypothesized that, if we amplified total DNA extracts from cells incubated with our nanostructures using primers specific for our scaffold sequence, we should observe template amplification in our reaction, suggesting that structures were taken up by cells, and can quantify the amount of nanostructures taken up using a standard curve.

The first step was to generate a standard curve for our scaffold, and we did so by creating a serial ten-fold dilution of our scaffold from 10^9 molecules (1.66 fmol) to 10^3 molecules (1.66 zmol). Forward and reverse primers reported by Okholm et al[33] for the M13mp18 genome were ordered (IDT, Coralville, IA). The template and primers were mixed with the Lightcycler 480 SYBR Green Master Mix (Roche, Basel, Switzerland) and amplified using the Bio-Rad CFX96 thermocycler (Bio-Rad, Hercules, CA). The resulting curves are shown in Figure 3.9.

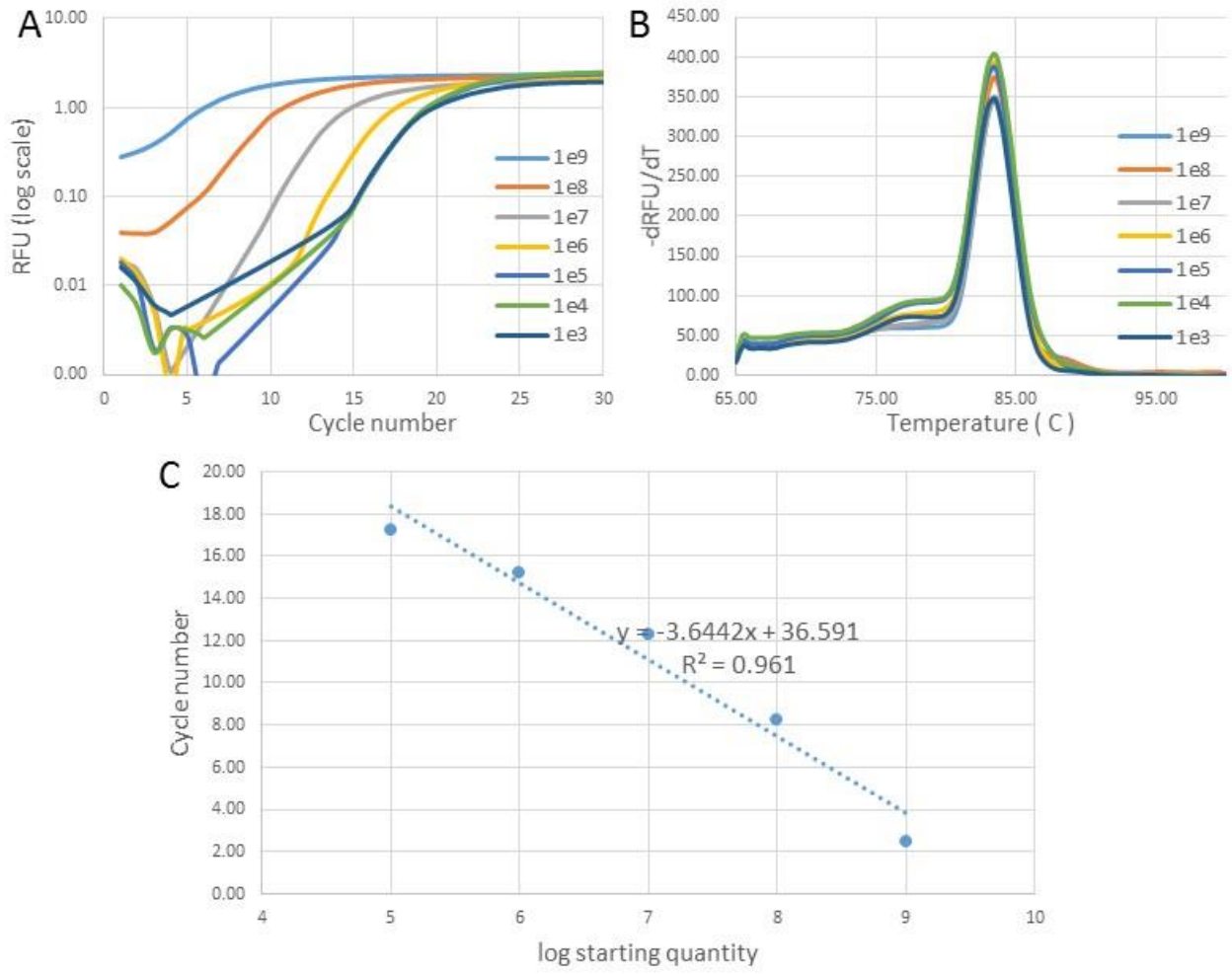

Figure 3. 9: (A) Amplification curves, (B) melt curves, and (C) standard curve for scaffold standards. Each reaction was performed in duplicate.

The melt curves (Figure 3.9B) for each of the standard showed the same behavior with a melting temperature of 83.5°C, suggesting that only the template was amplified. The amplification curves (Figure 3.9A) show distinguishable amplification curves for samples down to 10^5 molecules, suggesting that our reaction system can reliably detect down to 10^5 molecules in solution. Finally, the standard curve (Figure 3.9C) shows the expected downward trend with a high R^2 value. The efficiency of amplification can be calculated using equation 1.

$$E = 10^{-1/slope} \quad (1)$$

where 100% efficiency means that the number of template molecules doubled at each cycle. The efficiency for our reaction was 88.11%, which is lower than recommended, so more optimization must be done to achieve higher efficiencies.

The next step was to amplify the negative controls, namely the genomic DNA from cells and the PCR grade water supplied by the Master Mix. The genomic DNA was extracted from OSU CLL cells using the DNEasy Blood and Tissue Kit (Qiagen, Hilden, Germany). In both cases, there should not be any scaffold present in the sample, so no amplification was expected. However, as Figure 3.10 shows, in the PCR grade water, some DNA was still amplified. Furthermore, the melt curves show a single sharp peak with a melting temperature of 83.5°C, suggesting that the DNA amplified is very similar to the target DNA.

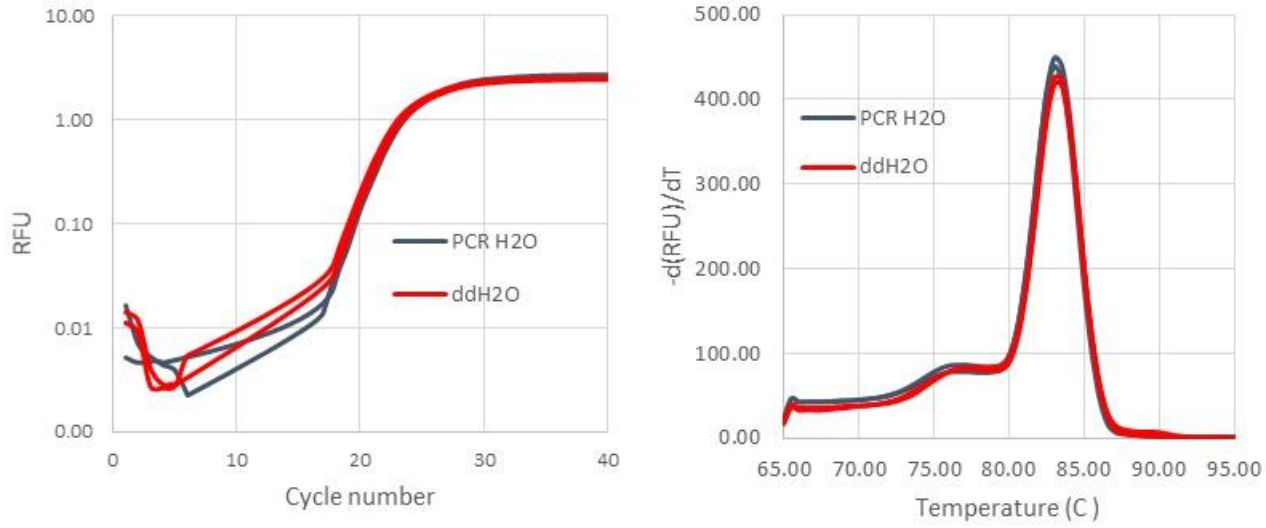

Figure 3. 10: Amplification curve (left) and melt curve (right) of double distilled water from the laboratory and PCR grade water provided in the qPCR kit.

This could be due to a few different reasons. The PCR grade water could have been contaminated by the items used to prepare the sample, such as tubes or pipette tips. Therefore, performing the reactions in sterile conditions could reduce the signal from the PCR sample. Indeed, amplification curves of reactions performed with PCR grade water under sterile conditions do show

slightly lower starting concentration of DNA, even though there is still amplification (Figure 3.11). However, the contamination in the PCR sample is such that it is outside of the detection range of our standard curve. Therefore, experiments can still be performed with that level of contamination.

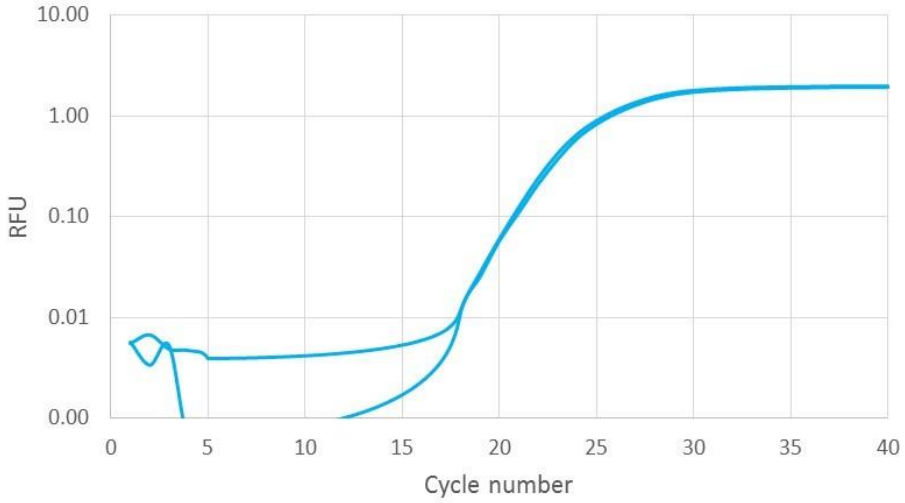

Figure 3. 11: Amplification curve of PCR grade water after sample was prepared under sterile conditions.

In terms of the genomic DNA, as Figure 3.12A shows, there seems to be non-specific amplification even in the absence of any nanostructure. Unlike the contamination of the PCR grade water, this contamination persists even under sterile conditions, suggesting that the primers are at least partially complementary to a region in the genomic DNA. The best way to remedy this would be to design new primers with the genome of the target cells in mind. However, in the meantime, the loss of resolution due to non-specific amplification of genomic DNA can be treated as a lower limit of this assay's detection capability. As Figure 3.12B shows, there seems to be about 4.5×10^5 molecules of nonspecific target in the genomic DNA. Therefore, using the current primers, about 10^6 molecules of target DNA can be detected reliably.

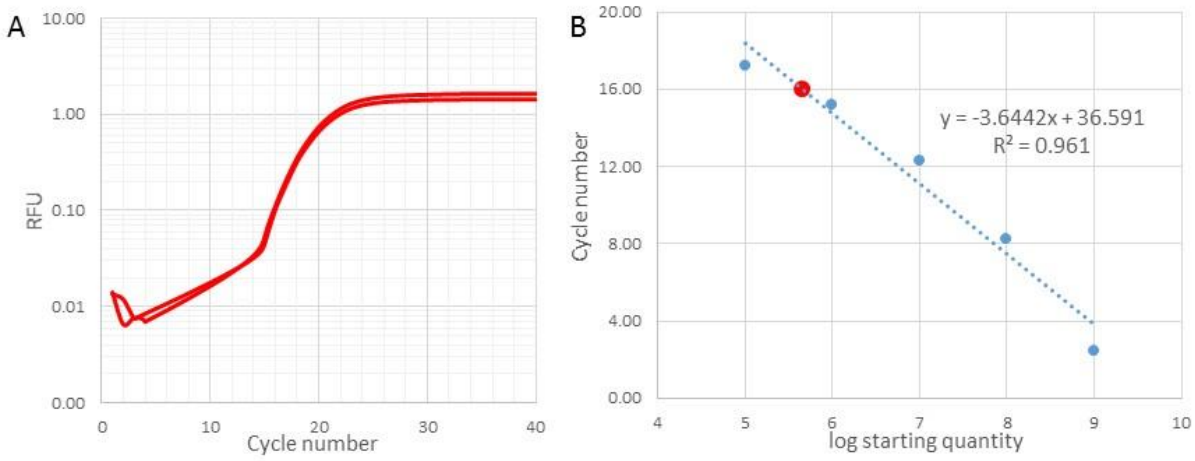

Figure 3. 12: (A) Amplification curves of genomic DNA from OSU CLL cells. (B) Standard curve from Figure 3.9 showing the starting quantity of nonspecific amplified DNA in genomic sample.

3.4. Sequestration in vitro

In the previous sections, the sequestration of miRNA in solution and the cellular entry of nanostructures were demonstrated. The next step was to show that structures were still functional within the cell by observing some downstream effect of miRNA sequestration *in vitro*. The was achieved using a Luciferase assay.

Luciferase is an enzyme which, among other things, is responsible for the glow of the fireflies. It catalyzes the breakdown of a set of compounds known as luciferin to produce light (Figure 3.13A). In an excess of luciferin, therefore, the intensity of light is directly proportional to the concentration of active luciferase enzymes in solution. This fact can be used to quantify various expression levels in cells by coupling it with the expression levels of luciferase.

Luciferase assays can also be used to determine the relative miRNA levels in cells using the schematic shown in Figure 3.13B. First, the luciferase gene must be modified to include the 3'UTR of a gene that is degraded by the miRNA of interest, thus ensuring that the miRNA will interact with the luciferase mRNA. Next, the luciferase gene must be transfected into the model cell. The baseline luminescence from the system can be obtained by transfecting cells with a copy of the luciferase that

does not contain the 3'UTR of the gene of interest. However, when the modified luciferase is transfected and expressed by cells, the miRNA will degrade some of the luciferase mRNA, and therefore the luminescence intensity will be decreased relative to the control.

Our assay was based on this same principle. We obtained two types of plasmids from the Byrd Lab, one containing the wildtype luciferase gene (-3'UTR), and the other containing the luciferase gene modified to include the 3'UTR of SHIP-1, a target of miR-155 (+3'UTR). We originally planned to use OSU CLL as our model cells, but due to the low transfection efficiencies associated with OSU CLL, we decided to use Human Embryonic Kidney (HEK-293T) cells instead. Cells were grown for 24 hrs in a 24-well culture plate before being transfected with the plasmid using Lipofectamine 2000 (Thermo Fisher Scientific, Coralville, IA). After 24 hrs, cells were incubated with nanostructure for 8 hrs. Afterwards, the Promega Dual Luciferase Assay kit (Madison, WI) was used to lyse and prepare the cell lysates for the luminescence readout per protocol. The luminescence intensity readout was recorded using Horiba FluoroMax-4 (Irvine, CA). Since bioluminescence was being measured, there was no excitation wavelength, and the receptor was set to receive signals from 300 nm to 650 nm.

Our controls were the luciferase signal intensity from cells transfected with the -3'UTR plasmid only, from cells transfected with the +3'UTR plasmid only, and from cells transfected with the +3'UTR and incubated with our scrambled control (scr) nanostructure (+3'UTR+scr). Our experimental data were from cells transfected with +3'UTR and incubated with the am155 nanostructure (Figure 3.13C). We hypothesized that the signal would decrease for the +3'UTR and +3'UTR+scr samples relative to the -3'UTR sample. However, the signal for the +3'UTR+m155 sample would be higher relative to the +3'UTR and +3'UTR+scr samples, and should approach the levels seen in -3'UTR.

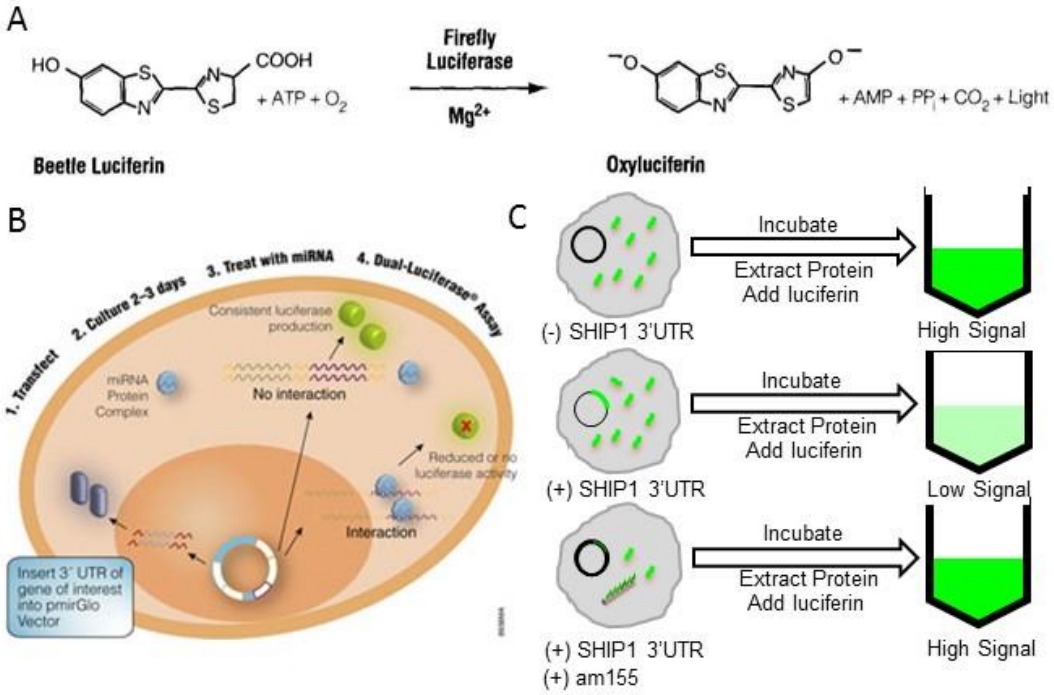

Figure 3. 13: (A) Chemical reaction showing degradation of luciferin by luciferase. (B) Schematic of luciferase assay used for quantifying mIRNA expression levels. (C) Schematic for quantifying miR-155 sequestration by am155 in vitro.

Figure 3.13C shows intensities at 527 nm for each sample tested. As expected, the signal intensity of the +3'UTR sample is lower than that of the -3'UTR control. Also, as expected, incubating cells with am155 nanostructures causes the signal to increase relative to the +3'UTR sample. However, there were a lot of discrepancies in the data. In terms of trends, the +3'UTR+scr sample should elicit a signal comparable to the +3'UTR control, but is more similar to the experimental sample. Furthermore, the decrease in signal of the +3'UTR is very modest, and could possibly fall within error. The graph shows data from only one trial, so the variability within the data is still unknown, and no conclusions can be drawn. However, the modest decrease in signal could possibly be attributed to the model cell, which may not have high expression levels of miR-155. The comparable signals between the am155 and scr samples could also be due to some effect that the structure itself, independent of the overhangs, has on luciferase expression. Further experiments must be conducted to answer these questions.

3.5. Conclusions and Future Work

We have successfully fabricated and characterized a nanostructure with ssDNA overhangs on its surface fully complementary to either miR-150 or miR-155. We have demonstrated selective sequestration of target miRNA in solution by these nanostructures with high efficiency. Furthermore, we have qualitatively shown cellular entry of these nanostructures without the need for any further modifications. We have established a qPCR protocol to create standard curves for DNA origami scaffold, and reduced possible contamination of samples using sterile technique. Finally, we have performed preliminary experiments to demonstrate sequestration of miRNA *in vitro*.

Future work will focus on quantifying cellular uptake and replicating the luciferase assay to better understand how structures behave *in vitro*. qPCR will be performed on experimental samples using the current protocol to assess its feasibility. New primers that may avoid nonspecific amplification of genomic DNA will simultaneously be designed. Finally, intensity analyses can be performed on fluorescent microscope images to also quantify the levels of structure uptake in cells.

For the luciferase assay, more controls, such as plasmids with a mutated 3'UTR, and native plasmids co-incubated with structures, will be developed and implemented to better isolate the effect of the functional overhangs on *in vitro* miRNA levels. In addition, different transfection protocols, such as electroporation, will be tested on OSU CLL cells in order to perform the assay in that cell line.

In the medium term, our structures must be tested for cytotoxicity, biodistribution, pharmacokinetics, and efficacy *in vivo*. If necessary, further modifications, such as poly-ethylene glycol (PEG) coating, 2'O-methyl RNA protection, or cell penetrating peptides may be implemented to improve structure performance. In the long term, our structures show great promise as an antisense drug delivery vehicle for any disease model with overexpressed miRNA. Furthermore, our work in antisense delivery can be synthesized with other areas such as chemotherapeutic delivery and antibody mediated

targeting to create a multifunctional drug delivery vehicle that is capable of attacking diseased cells on multiple fronts in a targeted manner.

4. Target Discovery

4.1. Overview

As stated earlier, the objective of this project was to develop a methodology to identify novel interaction partners to nucleic acids, and extract them from a complex mixture for analyses. As a proof of principle, we aimed to demonstrate successful isolation of ciRS-7 from cell lysate using a nanostructure presenting miD-7 overhangs on its surface. In this chapter, the D7 structure design will be elaborated on, the protocol for target discovery and isolation will be presented, and experimental results from the validation of each step of the protocol will be discussed.

4.2. D7 Structure Details

Figure 4.1A shows a model of the D7 nanostructure. The miD-7 sequences on the structure function similarly to the overhangs on the am150 and am155 structures, in that their task is to bind to their complementary strand, which in this case is the anti-miR-7 sequence. However, D7 also contains a special overhang (bioAdap) that is not related to any miRNA, but is complementary to a ssDNA strand with a biotin attached to it (bioDNA). Furthermore, as seen in Table 4.1, the overhang itself is 36 bases long, but the biotinylated strand is only 25 bases long. This is because the function of this overhang is to bind to the bioDNA until the bioDNA is displaced by a release strand (disp), which is fully complementary to the bioAdap, through toehold mediated strand displacement[34].

Strand name	**Sequence (5'->3')**
Biotin adapter overhang (bioAdap)	TTT TTT AAC CGC TCT GGT TAA CGT GTC TGG GCA AAA
Biotinylated strand (bioDNA)	TTT TGC CCA GAC ACG TTA ACC AGA G
Release strand (disp)	TTT TGC CCA GAC ACG TTA ACC AGA GCG GTT AAA AAA

Table 4. 1: Sequences of the different componenets of the biotin displacement mechanism. Sections with the same color have the same sequence in the case of bioDNA and disp, and are complementary in the case of either bioDNA or disp with bioAdap.

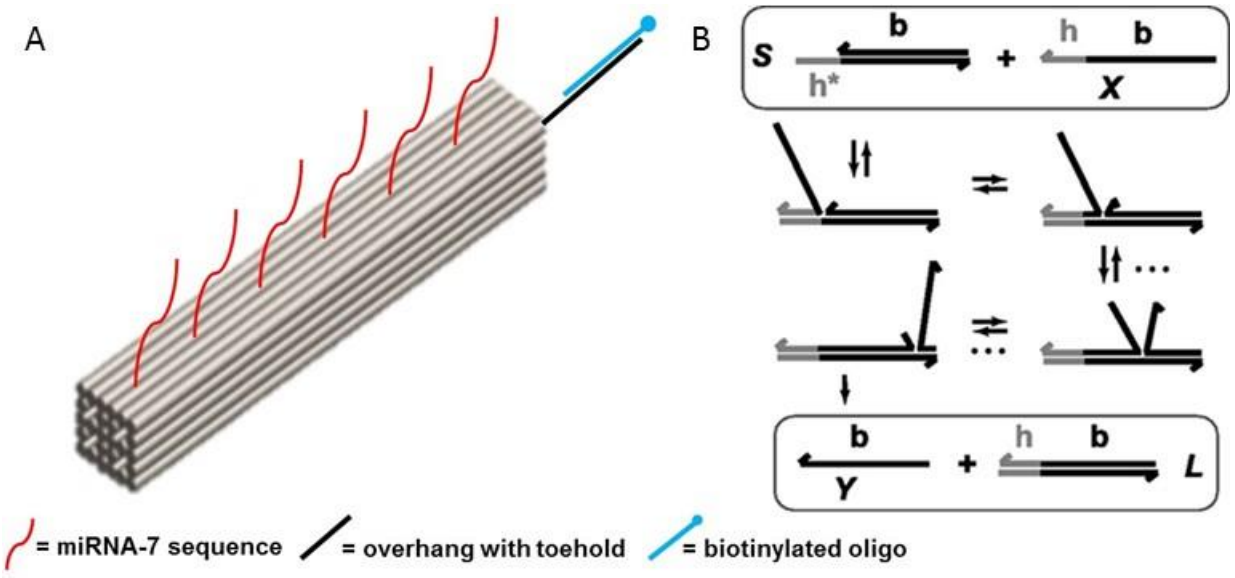

Figure 4. 1: (A) Model showing the different components of the D7 nanostructure. (B) Schematic showing the steps in toehold mediated strand displacement[34].

Toehold mediated strand displacement is a process by which an "invader" strand can displace an "incumbent" strand by being complementary to a single stranded domain on the "substrate" strand that the incumbent strand is bound to. This is shown in Figure 4.1B. In this case, X is the invader strand, and it is complementary to the entirety of the substrate strand, including to the h* domain, whereas the incumbent strand is not. The invader strand can bind reversibly to the single stranded "toehold" on the substrate, and on occasion, may compete with the incumbent strand in the b domain. The invader and incumbent strands may exchange base pairs back and forth, but ultimately the invader strand is more favored thermodynamically due to the greater number of base pairs, and will displace the incumbent strand[34].

In our structure, the 5' end of the bioAdap remains single stranded when the bioDNA is bound, whereas the disp strand is fully complementary to the bioAdap. Therefore, in the presence of disp in solution, the bioDNA is expected to be displaced completely by the disp strand. If the nanostructure was attached to another component through a streptavidin-biotin interaction, as will be the case, the displacement of the bioDNA will result in the release of the structure.

4.3. Process Schematic

The target discovery and extraction process was broken into three major steps, as shown in Figure 4.2.

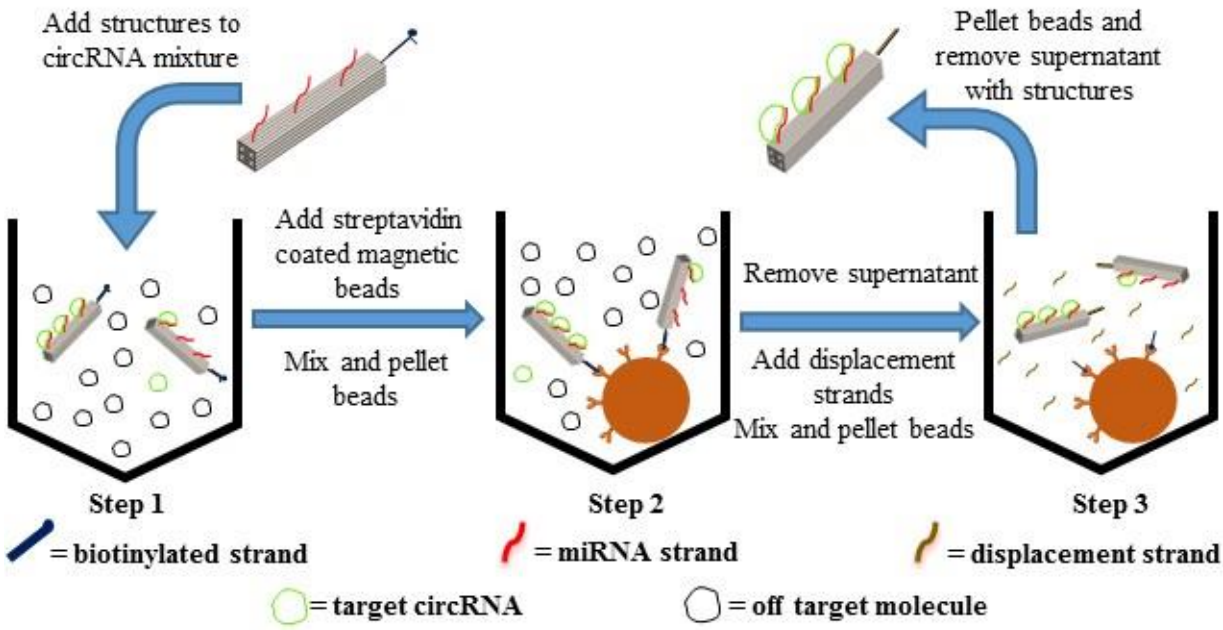

Figure 4. 2: Process schematic for target discover and extraction using nanostructures.

In step 1, the structure with the bioDNA attached (bioD7) is incubated a sample containing the target molecule and possible other off target molecules as well. The structure should be able to sequester the target molecule selectively from the solution, similar to miRNA sequestration in the previous project. In step 2, streptavidin coated magnetic beads are added to the solution and then mixed. During mixing, the structures should bind to the bead through streptavidin-biotin interactions. The interaction between the protein streptavidin and the small molecule biotin is one of the strongest known non-covalent interactions in biology, so the structures should stay attached to the beads. After mixing, the beads will be pelleted from the solution by applying a magnetic field to sample. This should remove a significant portion of the structures from the solution. In step 3, the supernatant, containing all the off-target molecules, will be removed, and the beads will be resuspended in the release buffer

containing disp strands and mixed. The disp strands will displace the bioDNA which will remain attached to the bead, and the structures will be released from the beads. The beads will then be pelleted and the supernatant containing the target molecule bound to the structure will be recovered for analysis.

4.4. Process Validation

Each step of the process was validated in isolation before the whole system was synthesized. During the validation, anti-miR-7 sequence conjugated with Cy5 fluorophore (far-7) was used as the target molecule for ease of detection.

For step 1, the successful incorporation of the bioAdap, the successful binding of the bioDNA, and the sequestration of far-7 were verified. The incorporation of bioAdap was tested by incubating D7 nanostructures with an ssDNA conjugated with ATTO647 fluorophore with the same sequence as the bioDNA. D7 structures were incubated with varying concentrations of the fluorescent ssDNA at 37°C for 2 hrs, then run on a gel, and imaged using the Typhoon gel imager. The results (Figure 4.3A), show that the fluorescence intensity does increase as more fluorescent ssDNA is present in the solution. However, the signal at a 1:1 ratio of fluorescent ssDNA to bioAdap was very low, suggesting that an excess of bioDNA is necessary for complete binding to the structure.

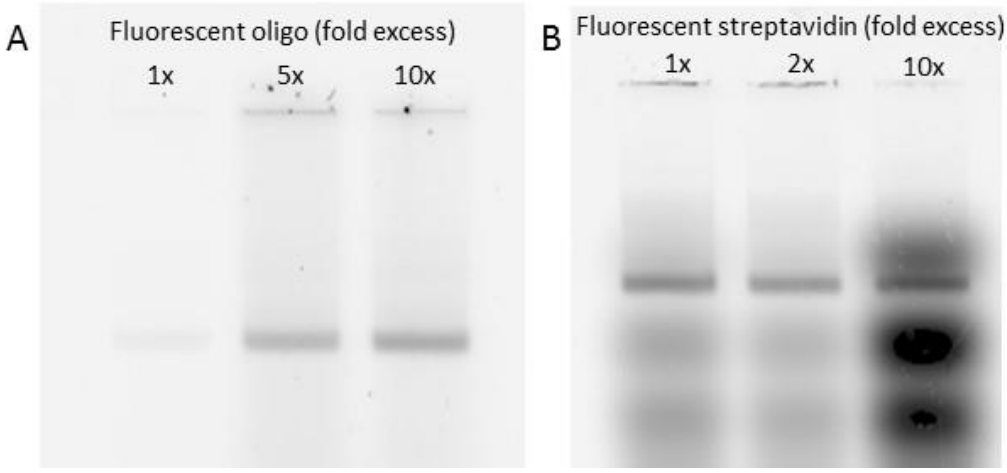

Figure 4. 3: (A) fluorescent gel assay showing successful incorporation of bioAdap into D7. (B) Fluorescent gel assay showing successful incorporation of bioDNA into D7. Both images were taken with 647 nm laser.

The incorporation of the bioDNA onto D7 to form bioD7 was also validated using fluorescent gel imaging, but with streptavidin conjugated with ATTO647. D7 nanostructures were incubated with 5x bioDNA for 2 hrs at 37°C. Afterwards, the nanostructures were purified using poly-ethylene glycol (PEG) precipitation following existing protocol[35]. The excess bioAdap had to be removed from solution so that the fluorescent streptavidin could only bind to biotin attached to the structures. The structures were then incubated with the fluorescent streptavidin at 37°C for 2 hrs, run on a gel, and imaged using the Typhoon. The results are shown in Figure 4.3B. The fluorescent signal from the band signifies that streptavidin successfully associated with the structure, suggesting that the bioAdap was incorporated into the structure.

The functionality of the miD-7 overhangs on the structure were tested using the same sequestration assay as for am150 and am155, except with only the target strand in solution. Therefore, it was expected that the assay would produce the same results. However, in the first few trials, the structures in the samples with over equal ratios of far-7 and miD-7 would aggregate in the wells and not run at all, thus causing the assay to fail. TEM images of those samples (Figure 4.4A) revealed that the structures were binding to each other in a distinct branching pattern, with the tip of one structure joined to the side of another. This branching led us to hypothesize that there might be some unwanted interactions between the miD-7 overhangs on the surface of the structures and the single stranded scaffold loops that were present at the ends of the structure. To test this hypothesis, we identified the sequence of the single stranded loops and compared them to the sequence of the D7 overhang sequence using MATLAB (code included in Appendix C) to identify the total number of complementary bases and number of sequential complementary bases. The analysis showed that there were multiple points on the single stranded scaffold loops with up to 6 sequential bases complementary to the overhangs, providing locations for possible binding and aggregation. The locations are denoted on the scaffold routing in Figure 4.4B. To remedy this situation, the staple routing was changed such that the

single stranded scaffold regions that were complementary to the overhangs were removed. The updated staple routing is included in Appendix A.

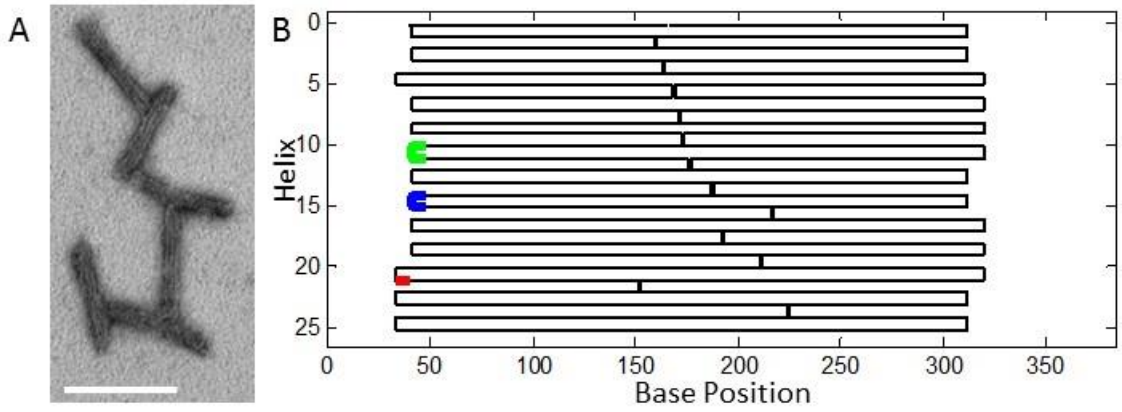

Figure 4. 4: (A) TEM image showing D7 nanostructures bound to one another in the branching pattern. Scale bar is 100 nm. (B) Scaffold routing with different regions of high sequential base complementarity highlighted.

The new version of D7 was then used to perform the same sequestration assay. The results, shown in Figure 4.5, shows that the sequestration assay works as expected, with no unexpected aggregation in the wells. This concluded the validation of the components of step 1.

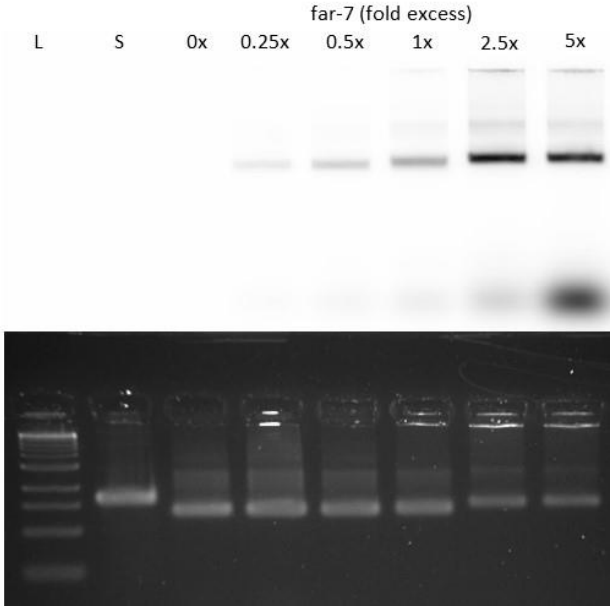

Figure 4. 5: Sequestration assay for far-7. The top gel shows image taken using the 647 nm laser. The bottom gel shows SYBR Gold total DNA stain.

In step 2, the binding of the structures to the streptavidin coated magnetic beads was validated. The beads were obtained from Spherotech (Lake Forest, IL) and had an average diameter of 4 um. First, the functionality of the beads were tested by incubating them with different concentrations of biotin conjugated with ATTO550 overnight. To prevent the beads from settling during incubation, the incubation was carried out on a rotor. After incubation, the beads were deposited on a glass slide and imaged using the fluorescent microscope. As Figure 4.6A shows, the fluorescence intensity around the bead increases after incubation in the fluorescent biotin compared to initial.

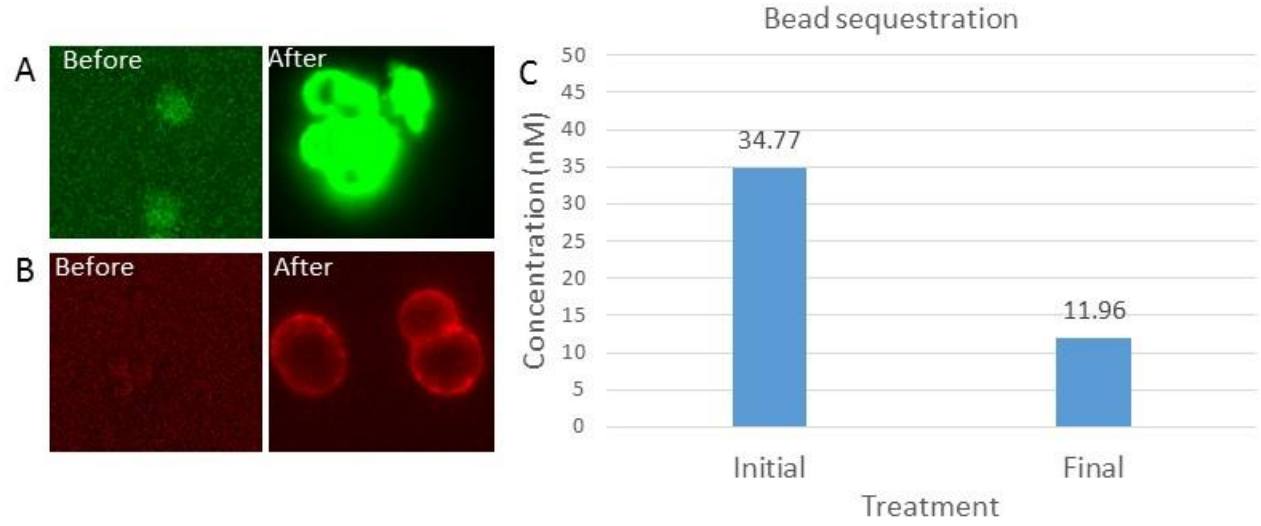

Figure 4. 6: (A) Fluorescent microscopy image of streptavidin coated magnetic beads imaged using the 532 nm laser before (left) and after (right) incubation with fluorescent biotin. (B) Fluorescent microscopy image of streptavidin coated magnetic beads imaged using 647 nm laser before (left) and after (right) incubation with far-7 bound bioD7. (C) Concentration of bioD7 in supernatant before and after incubation with streptavidin coated magnetic beads.

To test the attachment of nanostructures to the beads, bioD7 structures were first incubated with far-7 according to sequestration protocol and then purified through PEG precipitation. The fluorescent anti-miR-7 served as the signal for structure detection. The concentration of the prepared structures was measured on a Nanodrop spectrophotometer (Thermo Fisher Scientific, Coralville, IA) before being incubated with the magnetic beads on a rotor overnight. After incubation, the beads were pelleted and the concentration of nanostructures were measured again using the Nanodrop. In addition,

the beads were also imaged using the fluorescent microscope before and after incubation with the nanostructures.

As seen in Figure 4.6B, the fluorescence intensity around the magnetic beads increased after the incubation with bioD7 with far-7 bound to it, suggesting that the beads can sequester the structures from solution. This effect was further quantified using Nanodrop measurements, which shows a drop in concentration of the nanostructures in the supernatant after incubation, suggesting that some structures are bound to the beads and is therefore not freely available in the solution. However, even after an overnight incubation, there was still over 10 nM of structure in the supernatant, suggesting that the efficiency of binding to the beads may be low.

Finally, in step 3, the dissociation of the nanostructures from the beads through strand displacement was tested. First the displacement of bioAdap bound to free streptavidin was studied. To accomplish this, bioD7 structures were incubated with ATTO647-conjugated streptavidin at 37°C for 2 hours, and purified using PEG precipitation. The sample was then incubated with various excess concentrations of disp staple before being run on a gel and imaged using the Typhoon. Initially, excesses of 5 fold and 10 fold were studied, similar to the other assays. However, no displacement was observed at those concentrations. Therefore, to increase the probability of a successful displacement, the excess concentration was increased to 50x and 100x, and the incubation temperature was increased to 40°C. Under those conditions, strand displacement was observed, as seen in Figure 4.7. Studies have been published showing possible interactions between streptavidin and adenine nucleotides[36], and there are multiple adenines proximal to the streptavidin in our system, so we hypothesize that the possible additional interactions could have required more energy to be supplied to the system for successful displacement.

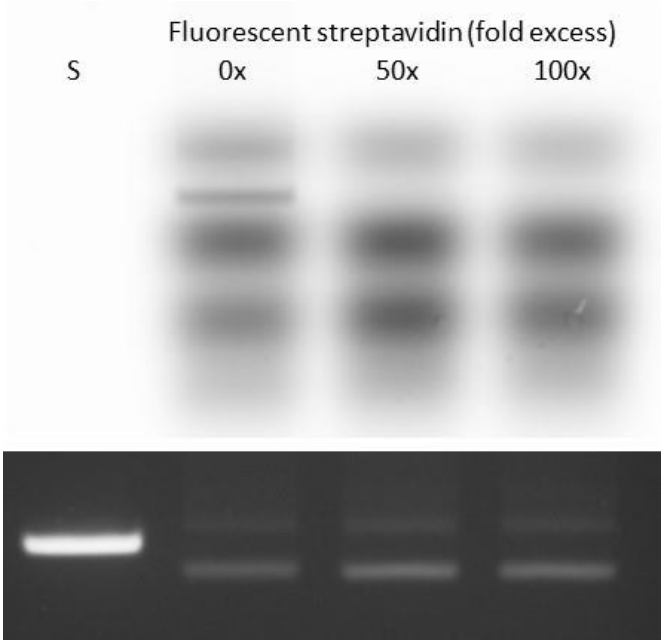

Figure 4. 7: Fluorescent gel assay showing successful displacement of fluorescent streptavidin associated bioAdap strand. The top image was obtained using the 647 nm laser, and the bottom image was obtained using SYBR gold total DNA stain.

Lastly, the dissociation of the structures from magnetic beads was tested. This was done by analyzing the supernatants obtained after different steps of the process through fluorescent gel imaging and total DNA staining. The nanostructures used were bioD7 with far-7 bound as an indicator. The first sample (post-seq) was taken after the structures were incubated with the beads. Since it was shown that the bead sequestration is not very efficient, we expected some structure to be present in the post-seq sample. However, this also means that any sample taken after the release solution is added will potentially be contaminated with structures that were never attached to the bead. Therefore, a series of washes were conducted before the release step, and the supernatant from the washes was analyzed to see whether any structure remained. After *n* washes, the release solution containing disp strands was added, and the resulting supernatant was analyzed. We hypothesized that the structure signal would fade after the wash, and would recover after addition of the release solution. The schematic of this process is shown in Figure 4.8.

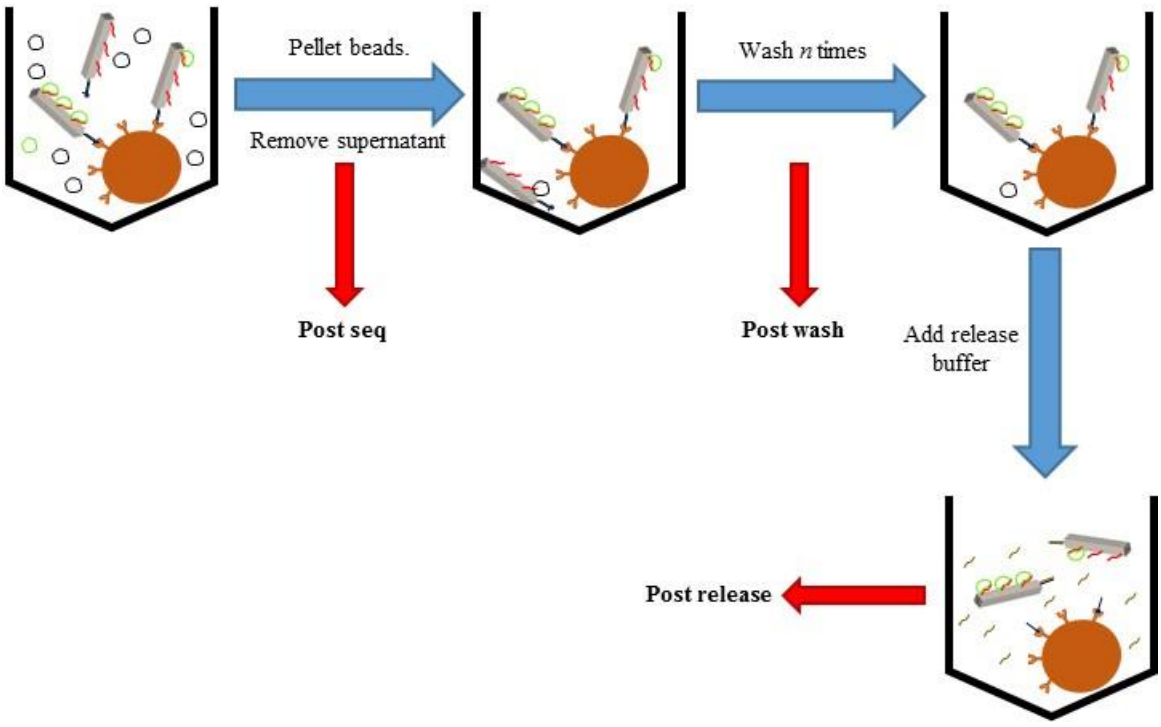

Figure 4. 8: Schematic for the verification of structure release from beads.

The results of this experiment is shown in Figure 4.9. 500 µL of 5 nM bioD7 was incubated with the magnetic beads on a rotor overnight. Afterwards, the beads were pelleted and the supernatant was stored for analysis. The beads were then washed three times using 200 µL of 1x FOB with 18 mM $MgCl_2$, the storage buffer for the structures. The supernatant from each wash was collected for analysis. Finally, the release solution containing 100 fold excess displacement staples were added to the beads, and the beads were shaken and incubated at 40°C for 4 hrs. After the incubation, the beads were pelleted and the supernatant was collected for analysis.

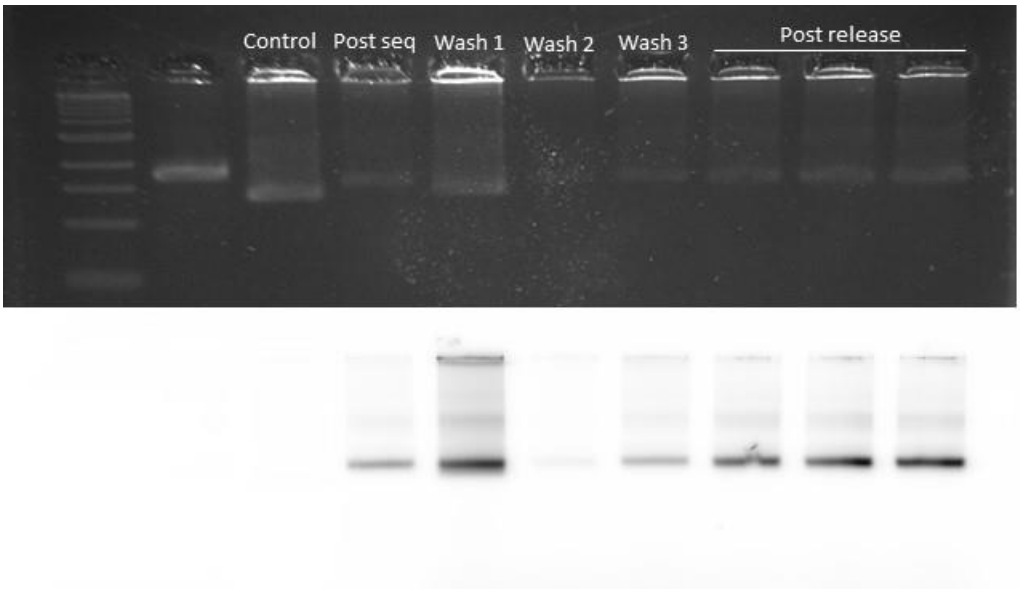

Figure 4. 9: Results of the bead release assay. The top gel shows total DNA stain using SYBR Gold, and the bottom gel shows fluorescent gel image using the 647 nm laser. The three post release wells are replicates of the same sample.

As expected, the post-seq supernatant still contained some structures, showing that the bead sequestration is not highly efficient. The fluorescence signal from the band jumped after the first wash, which was unexpected. This could be due to the wash volume being lower than the initial volume (200 µL as opposed to 500 µL), so the remaining structures could have gotten more concentrated. After wash 2, the structure signal seems to have mostly faded from both the total DNA and fluorescent gels, suggesting that washes are successful at removing excess structure. However, the signal seemed to return after wash 3. The force of the multiple washes could have dissociated some structures from the beads, which could be causing this return in signal. As expected, though, the structure signal did recover after the addition of the release solution, both in the total DNA stain and in the fluorescent channel, suggesting that the strand displacement reaction did occur. It must be noted, however, that this gel represents only one replicate of the experiment, so the discrepancies observed in the gel may be due to random error. More trials must be performed to draw conclusions.

4.5. Conclusion and Future Work

We have successfully designed and fabricated a nanostructure with miD-7 sequences on the overhangs, and with a biotin adapter system that can bind to a biotinylated strand and subsequently remove it using toehold mediated strand displacement. We have developed a protocol to use these structures to select and extract any molecule that interacts with the sequence present on the overhangs. As a proof of principle, we have validated all the steps of the protocol individually using fluorescently labeled anti-miR-7.

Future work in the near term will include synthesizing and testing the entire process in solution, testing the sequestration steps using ciRS-7, and in harsher environments such as cell lysate. A model system for the proof of principle must also be identified to test the protocol *in vitro* on naturally expressed circRNA.

The field of circRNA is very new and in constant flux. The publication by Guo et al[27] challenged earlier notions that many circRNA act as miRNA sponges, and that claim still stands. However, most of these conclusions are drawn using bioinformatics tools, and an empirical method such as ours may be able to discover interactions that do not satisfy the assumptions of these techniques. Therefore, the methodology may still be used to detect for circRNA in the future.

Perhaps more importantly, the general methodology is not defined by the nature of the target molecule. In theory, it should be able to identify and extract any interacting partner to nucleic acids, including proteins[31]. Therefore, the focus of this methodology could shift to different classes of interacting partners, providing a new avenue to explore.

Works Cited

1. Pray, L. “Discovery of DNA Structure and Function: Watson and Crick.” Nature Education 1.1:100 (2008).
2. Seeman, N. “Nucleic Acid Junctions and Lattices.” Journal of Theoretical Biology 99 (1982):237-247.
3. Chen, J., Seeman, N. “Synthesis from DNA of a molecule with the connectivity of a cube.” Nature. 350 (1991):631-633.
4. Seeman, N. C. & Lukeman, P. S. Nucleic acid nanostructures: bottom-up control of geometry on the nanoscale. Rep. Prog. Phys. 68, 237–-270 (2005).
5. Rothemund, Paul. “Folding DNA to create nanoscale shapes and patterns.” *Nature* 440 (2006):297-302
6. Douglas, Shawn M. et al. “Self-assembly of DNA into nanoscale three-dimensional shapes.” *Nature* 459 (2009):414-418.
7. Dietz, Hendrik, Douglas, Shawn M., & Shih, William, M. “Folding DNA into twisted and curved nanoscale shapes.” *Science* 325 (2009):725-730.
8. Marras, Alexander E. et al. “Programmable motion of DNA origami machines.” *PNAS* 112 (2014): 713-718.
9. Ketterer, Phillip, Willner, Elena M. & Dietz, Hendrik. “Nanoscale rotary apparatus formed from tight-fitting 3D DNA components.” *Science Advances* 2 (2016).
10. Douglas, Shawn M., Bachelet, Ido, Church, George M. “A logic gated DNA nanorobot for targeted transport of molecular payloads.” *Science* 335 (2012):831-834.
11. Jiang, Qiao et al. “DNA origami as a carrier for circumvention of drug resistance.” *JACS* 134 (2012):13396-13403.
12. Zhang, Qian et al. “DNA origami as an *in vivo* drug delivery vehicle for cancer therapy.” *ACS Nano* 8 (2014):6633-6643.
13. Perrault, Steven D. & Shih, William M. “Virus-inspired membrane encapsulation of DNA nanostructure to achieve *in vivo* stability.” *ACS Nano* 8 (2014):5132-5140.
14. Douglas, Shawn M. et al. “Rapid prototyping of 3D DNA-origami shapes with caDNAno.” *Nucleic Acids Research* 37 (2009):5001-5006.
15. Castro, Carlos E. et al. “A primer to scaffolded DNA origami.” *Nature Methods* 8 (2011):221-229.
16. Sobczak, Jean-Philippe J. et al. “Rapid folding of DNA into nanoscale shapes at constant temperature.” *Science* 338 (2012):1458-1461.
17. Biochem text book
18. Wightman, B., Ha, I. & Ruvkan, G. “Posttranscriptional regulation of the heterochronic gene lin-14 by lin-14 mediates temporal pattern formation in *C. elegans*.” *Cell* 75 (1993):855-862.
19. Almeida, M.I., Reis & R.M., Calin, G.A. “MicroRNA history: discovery, recent applications, and new frontiers.” *Mutat Res* 717 (2011):1-8.
20. Siomi, H. & Siomi, M.C. “Posttranslational regulation of microRNA biogenesis in animals.” *Mol. Cell* 38 (2010):323-332.
21. Fulci, Valerio et al. “Quantitative technologies establish a novel microRNA profile of chronic lymphocytic leukemia.” *Blood* 109 (2007):4944-4951.
22. Ferrajoli, Alessandra et al. “Prognostic value of miR-155 in individuals with monoclonal B-cell lymphocytosis and patients with B chronic lymphocytic leukemia.” *Blood* 122 (2013):1891-1899.
23. O’Connell, Ryan M. et al. “Inositol phosphatase SHIP1 is a primary target of miR-155.” *PNAS* 106 (2009):7113-7118.
24. Mraz, Marek et al. “miR-150 influences B-cell receptor signaling in chronic lymphocytic leukemia by regulating expression of GAB1 and FOXP1.” *Blood* 124 (2014):84-95.
25. Yeh, Yuh-Ying et al. “Characterization of CLL exosomes reveals a distinct microRNA signature and enhanced secretion by activation of BCR signaling.” *Blood* 125 (2015):3297-3305.
26. Wang, Peter L. et al. “Circular RNA is expressed across the eukaryotic tree of life.” *PLoS ONE* 9 (2014):1-10.
27. Guo, Junjie U. et al. “Expanded identification and characterization of mammalian circular RNAs.” *Genome Biology* 15 (2014):1-14.

28. Memczak, Sebastian et al. “Circular RNAs are a large class of animal RNAs with regulatory potency.” *Nature* 495(2013):333-338.
29. Hansen, Thomas B., Kjems, Jørgen & Damgaard, Christian K. “Circular RNA and miR-7 in cancer.” *Cancer Research* 73 (2013):5609-5612.
30. Iorio, Marilena V., Croce, Carlo M. “MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review.” *EMBO Molecular Medicine* 4 (2012):143-159.
31. Eiring, A.M. et al. “miR-328 functions as an RNA decoy to modulate hnRNP E2 regulation of mRNA translation in leukemic blasts.” *Cell* 140 (2010):652-665.
32. Halley, Patrick D. et al. “Daunorubicin-loaded DNA origami nanostructures circumvent drug-resistant mechanisms in a leukemia model.” *Small* (2015):308-320.
33. Okholm, Anders H. et al. “Quantification of cellular uptake of DNA nanostructures by qPCR.” *Methods* 67 (2014):193-197.
34. Srinivas, Niranjan et al. “On the biophysics and kinetics of toehold-mediated DNA strand displacement.” *Nuc. Acids. Res.* 41 (2013):10641-10658.
35. Stahl, Evi et al. “Facile and scalable preparation of pure and dense DNA origami solutions.” *Angew. Chem.* 53 (2014):12735-12740.
36. Bing, Tao et al. “Specific interactions between adenosine and streptavidin/avidin.” *Bioorg. Med. Chem. Lett.* 22 (2012):7052-7055.

Appendix A

Table A 1: Staple sequences for structure fabrication.

Sequence Name	Sequence
Core_01_01	CTG GCA TGG CGA GAA ATG TAA ATC TTA GTC TT
Core_01_02	AGC TGG CGG GCC AGT GCC TTC CTG TAG CCA GC
Core_01_03	TGG GAA GGG CGA TCG GAC TCT AGA GGA TCC CC
Core_01_04	GCC CGA ACT GAT GGC AAT TCA TCA ACA GTA CC
Core_01_05	ATG CCT GAG TAA TGT GTA GGT AAA CCC CGG TT
Core_01_06	CAT ATC AAA ATT AAA ATC ATA CAG GGG TAG CT
Core_01_07	GCG AAA GGC GAT TTA GAG CTT GAC CCA TAT TT
Core_01_08	CGG CCA ACA CTC GCC AGG GTT TTC TGT GAG CG
Core_01_09	TTG CCC CAA AAT CGG CAA AAT CCC ACA TTC AA
Core_01_10	TTT AGC GAA CCT CCC GAC TTG CGG GGG TAT TA
Core_01_11	AAC GAG TAG TTA CTT AGC CGG AAC TTG TAT CA
Core_01_12	GTT GGT GTA GAT GGG CTT AAC CAA TAG GAA CGC C
Core_01_13	ATA CGA GCT CCT GTG TGC CGG AAA CAG TAT CGA AGC
Core_01_14	GGG TGC CTA GCT CGA ACA GGC TGC ACC GTG CAC CCG
Core_01_15	AAG ATG ATT AGC GAT AAC ATT CAA AAA TAC ATG GTA
Core_01_16	ATA GTA TGG TTT GAA ACG CAG AAT TAA CGA GAG AAT
Core_01_17	AAA GAA CAT GTT TTA AAT ATC GCA AAT GCC TGT AAT
Core_01_18	GCC GTT TTC AAC ATG TAC AAA TTC TTA CCA GTA ACC
Core_01_19	GGA TGT CTG GAA GTT TCA TTG AGT AGA TTC AAC GCA
Core_01_20	ATT GTA ACG TCA AAA ATG AAT AAG AAC GAC AAG CAA
Core_01_21	AAT GAC AGC GCC ATT CGC CAT TTT CGT AAT ATT TTT GT
Core_01_22	GAG CCA CCT TGC TTT CTC ACC GTC AAT AAG AGC AAA AGA A
Core_01_23	TCT GAA TTT ACC GTT CTC ATC GGC ATT TTC GGT CAT AGC C
Core_01_24	GCC ACC CTA AAG GAG CCT TTA ATT CTA TCT TAA AAC CGA G
Core_01_25	GCG GAT AAG AGC CGC CGC CAG CAT TGA CAG GAT CAG AGC C
Core_01_26	TAG CCG AAC GTC CAA TAA GCA AAC AAC GGC GGA TTG ACC G
Core_01_27	CAC TGA GTT TCG TCA CAG ACA GCC CTC ATA GTA ACA ACT A
Core_01_28	CCA TCT TTT GGG AAT TAG AGC CAG AGA TAA CCA CGC AGT A
Core_01_29	TAA TGC GCA TAC CGA AGA GTA GAA GAA CTC AAT ACT ATG G
Core_01_30	ATC AAT ATT TTT AAA AGT TTG AGT AAC ATT ATG CGG AAT T
Core_01_31	AAG GCT ATT TTC ATC AAT TCT CCG ATG TGC TGT AAT TAG A
Core_01_32	TTA AGT TGG GTA AAT CAG CCC TAA TAT GTG AGA TGG AAA C
Core_01_33	AGA CTT TAC AAA CAA TTT GGA TTA TAC TTC TGT TTC AGG T
Core_01_34	GAC GCT CAA AAG AGT CCC ACG CTG AGA GCC AGA ACC CTC A
Core_01_35	TAG CTG TTC GGA AGC ATA AAG TGT AGC GGT CCA CGC TGG T
Core_01_36	GAT TAG TAT GCT TTC CTC GTT AGA CGC GTA ACC ACC ACA C
Core_01_37	ACC ATT AGA GAA CCG CCA CCC TCA TTC AGG GAT AGC AAG C

Core_01_38	AGG GTG AGA AAG GCC GCC AAA AAC ATT ATG ACG TCA ATA A
Core_01_39	AGT ACA TAA ATT TTG CAC GTA AAA CAG AAA TAA GAG AAG G
Core_01_40	GTA TTA ACT CAA CAG TTG AAA GGA TCG TAT TAA ATC CTT T
Core_01_41	ACG CGA TAG AAC CCT TCT GAC AAC AGT GTG TCC ATC ACG C
Core_01_42	ATT AGG ATC AGG TCA GAC GAT TGG CCT TGA TAC GGA ACC A
Core_01_43	CTG GAT AGC AAA GTT ACG ACA ATG AGG CTC CAC AGA ACC G
Core_01_44	TTG CTT TGG CCG CTA CAG GGC GCG GCG CGG GGC AGG GCG A
Core_01_45	GAA ACG CAA TTT CAT CAA TCA ATA AGT AAC AAC CCG TCG G
Core_01_46	AAC AAC GCT ATT TTC ATC GTA GGA ATA TAG AAG GCT TAT C
Core_01_47	CCG ACA AAT TAT CAT TCC AAG AAC GAG GTT TTG AAG CCT T
Core_01_48	AAC GGA ACA AAT CTA CGT TAA TAA AAC CGG ATA TTC ATT A
Core_02_01	CTA ATG CAT TGT GAA TTA CCT TAT TGA CGA GAA ACA CCA G
Core_02_02	AAC CAA GTC CAG TAA TAG AAA AAG CCT GTT TAA TGT AAA T
Core_02_03	TGA ACA AGT GTT TAT CGA GGG GGT AAT AGT AAA ATG TAT T
Core_02_04	ATG GCT TAA AGC GAA CAC GCA TAA CTA AAA CGA AAG AAG A
Core_02_05	CAC CCA GCT ACA ATT TTT TCC AGA GCC TAA TTA ATG AAA T
Core_02_06	TTC ATC AAG AGT AAT CCA TAG GCT GGC TGA CCC TTT TTA A
Core_02_07	TTG GGC GCC AGG GTG GAC TCC AAC GTC AAA GGT AAT GCA G
Core_02_08	TTC AGC GGA AAG TAC ATG AGG AAG TTT CCA TTA GAA AAC GAG
Core_02_09	AGA GCC ACC ACC CTC ATC ACG TTG AAA ATC TCA GCG GCA AAA GA
Core_02_10	CAT TAG CAA GTC AGA GAC ATA AAG GTG GCA ACG CTT AGG TTA AG
Core_02_11	CGC GAA ACA GTG AGA ACC GCT TTT GCG GGA TCA AAG ATT AAG AG
Core_02_12	TGG CCC ACT CCA GAC GAC GAC AAT GTA GAA ACC AAT CAA TCT ATT TTG
Core_02_13	AAG GCG TTT AAT TCT GTA CGT GAA CCA TCA CCC TTA ATG CAC GAG CAC
Core_02_14	CGC CTC CCG GTT GAG GTA GCG GGG TTT TGC TCT AAC ATC CAC CTT TAA
Core_02_15	GGA AAC CTA AAA CGA CAA AGG GGG TGG GAA CAT CCA ACA GTT TGA TAA
Core_02_16	CAT TCC ACC AGT ACA ACG GAA TAG GTG TAT CAG CTA TAT TTA TAA TGC
Core_02_17	CGA TCT AAA ACC CAT GTA GTA CCG CCA CCC TCA TAC ATT TGC AAC TAA
Core_02_18	CCA CCC TCC ACC ACC AGT GCC GTC GAG AGG GTG GTG GCA TTT TTG CGG
Core_02_19	CCC TTA TTA ATG GAA AAA CCT ATT ATT CTG AAA GAT GAA TAA ACA AAA
Core_02_20	AAA ATC ACT TCA CAA AAG GCT GAG ACT CCT CAA AGA AAT TTG AAT TAC
Core_02_21	ACA TTA AAC CAG TCA CGA CGT TGT GTC GTG CCA GCT GCA TTT TGC GTA
Core_02_22	AAA TAT CAC AGC AAA TCA TTC TGG AAA TGG ATA CCT TTT TAT AAA GCC
Core_02_23	TTG GCA AAA CCG CCT GCC TGA AAG GAA ATA CCA TAA CTA TGT ATC ATA
Core_02_24	GCA AAC AAA ATA ATT CTA ATG GGA TCT TCG CTA TCG TCA TAA AAG AAG
Core_02_25	AAA CTA GCC TCA TTT TGC ATC GTA GCA ACT GTA TGC TTT ACG ATA AAA
Core_02_26	GAT AAT CAT TAA AAT TGA CGA CGA CCA GGC AAC ATA AAT CCA ACA CTA
Core_02_27	AAC TAA TAC ATT AAA AGA ACT GAT GGC CTT GCA AAT ATA TTG ATA AAT
Core_02_28	GCT GAA AAT GAT ATA AGT ATA GCC ACT ACA ACG AGC CGC CAC CAG AAC
Core_02_29	GCC GGA GAG CAA GGC AAA GAA TTA CTA ATA GTA GTA GCA TAG TAC CAG
Core_02_30	ATT ACC TGA CAA TAA CAT TCC TGA TTA TCA GAG TTA TTA ACT GGT CAG
Core_02_31	CAT CAA GAA TAC AGT AAT ATA ATC CTG ATT GTT CGA CAA CAT TGA GGA
Core_02_32	TTA ATT ACT AAT TAA TAT TGA CGG TCC TTA TTC ACA AGA AAT CCC AAT

Core_02_33	ATT TCA TTG CGT AGA TAA TAA TGG AAG GGT TAA GGA TTT AGC ACT AAC
Core_02_34	TTG CTC CTG TCA GGA TAG TTG CGC CCA GAA GGC CGA AGC CAA CGC TAA
Core_02_35	GAG GTC ATC AAT TCT AGC AAA ATT AAG CAA TAA ACC GTT CAG TCT GGA
Core_02_36	CTT TTT TAT GAA TAA CGA GGT GAA GGA ATA CCC AAG AAA CTG CCA GTT
Core_02_37	TGT AGC TCA CTT CAA ATG CAG GGA AAA ATA CGC AGC GAT TAC TGA CCA
Core_02_38	AGT ACG GTT GCA TCA AGT CAC CCT CTT TTT CAA CGG AGA TGA GGC GCA
Core_02_39	TTG CGT TGG CAG GTC GTG CGG GCC TAG GTC ACG AGC TTC AGA GCT TAA
Core_02_40	GCG CGT TTC AGT AAG CTA TAA ACA GTT AAT GCC GGG AGA AAG CAA AAG
Core_02_41	ACT CAT CTG CGA ATA ATT CGG TCG CTG AGG CTT ATC GCG TCC CCT CAA
Core_02_42	TGC GTT ATA ATT TAG GCC CTA AAG GGA GCC CCA GCG GGC GAA GGG ATT
Core_02_43	ATA AAC ACC AAA GAA CAT TAA GAC AAA TTA TTG AGC CAT TTC ATA ATC
Core_02_44	CGA CCG TGT TTA GTT AAT AAT AAC TTT CTT AAT TAT CAG CAC CGG AAC
Core_02_45	TAA TGG TTT GTT CAG CGC GAA AAA CCG TCT ATA GAG GCG GTA ATG AAT
Core_02_46	TTT TGC CAA ACA ATA GCT ATT AAA GAA CGT GGT TTT TCT TTC CAG TCG
Core_02_47	GCT TTT GCA AAT ATT CGC ACC AAC CCG ATA TAT AAT TTT TGC CTG TAG
Core_02_48	ACC AAA ATT ACA GGT ATG AGT GTT GTT CCA GTT GAT TGC CCA CAT TAA
Core_03_01	CCA GAC GAA ACA GTT CAA ACG GGT GTT AAA GGT AGA AAG GTA GCG TAA
Core_03_02	TCA TAA CCG GAA TAC CTT ATA AAT CAA AAG AAT TGC AGC AAA AGC CTG
Core_03_03	AGT AAG AGA AAA ATC AAC TAA AGA CAG CAG CGA ACA ACT TCA GAC GTT
Core_03_04	GTA TAA CGA TAA CAT CAA TAT TAC AAT GGC TAG TCG CTA TAT TTA ACA
Core_03_05	GGG AGC TAT AAC CGT TCG CTC ATG CGT AAG AAT GAA AAC AGA AAC AAA
Core_03_06	TTA GAC AGA CCG AGT AAT CGT CTG CCA ACA GAT GAG AAG ATC ATT TCA
Core_03_07	TTG CTG AAT TCA TTT GAG CTA AAT CGG TTG TAG AGA CAG TTA ATC GTA
Core_03_08	ACA AAA TAA TTA GTT GAA TCG GCT GTC TTT CCA GGT AAA GAA ATA AGA
Core_03_09	CGA GCG TCT ATC CTG ACC TAA TTT ACG AGC ATA AAC AAC ATG AAA TAC
Core_03_10	ACT TTG AAA ACG TAA CAG GAC GTT GGG AAG AAA ACA TTA TAG CGA GAG
Core_03_11	GAC GGT CAG CTT GCC CGC GAT TTT AAG AAC TGG AGA TTT ACT CGT TTA
Core_03_12	AAG GAA TTT TGA CCC CTA ATG CCA CTA CGA AGA TTG AAT CTT TAA TTC
Core_03_13	TAT CAG AGC AAA ATC AGG GAG GGA AGG TAA ATT TTC CCT TAA ATC CAA
Core_03_14	TGT TAG CAG CTG ATG CAG AAT CCT TAC GTG GCA CAG ACA AAG AGG TGA ACT TCT TT
Core_03_15	GCG TCT GGC CAA GCT TGC ATG CCT CGC TCA CTG CCC GCT TTT CAC CAG TGA GAC GG
Core_03_16	TCC AGA ACA CTT GCC TCG AAC CAC CAG CAG AAT AAA ATA TCT TTA GGA GAA GTA TT
Core_03_17	ATC ATC ATG GAT TCG CCT GAT TGC TGA GTA ACA GTG CCC GGT CAT ACA TGG CTT TT
Core_03_18	TTA ACG TCA CAT GAA AGT ATT AAG CAA ATA AAT CCT CAT TAA AGC CAG AGC GTT TG
Core_03_19	TAC CAG TCA AAG CTG CTC ATT CAG TGA ATA AGA TCA TAA GGG AAC CGA ATA CCA AG
Core_03_20	CCT GGC CCT GAG AGA GTA GCC CGA GAT AGG GTG AAA GAT TCA TCA GTT GCT CAT TA
Core_03_21	CTG AAC AAA GGC CGG ACA AAG ACA AAA GGG CGG CTT AGA TTG GGT TAT TAC ATT TT
Core_03_22	AGC CCA ATA CCG ACT TCA TTA AAG GTG AAT TAC TTG CTT CAC TTT TTC TGG TAA TA
Core_03_23	AGC AAT AGG TAT CGG TAC AGC TTG ATA CCG ATT AGA GAG TTT CTG ACC CAA GGC GA
Core_03_24	GAA AAG TAC AAA AAA AAC AAC AAC CAT CGC CCC AGA CCG GAC TGC GGA ATT ACG CC
Core_03_25	AGG TTA TCG ATA AAA CTA TTT TTG CGC CAG CCT CGC AAG ACG GAA TCA TAA TTA CTA AG
Neighbor_01	CCA TCG ATA GCA GCA CAT CAA TAG
Neighbor_02	AGA CGG GAA AGA CAC CAC GGA ATA

Neighbor_03	ATT TCA ACT TTA ATC AGA TAC ATA CGA GGC AT
Neighbor_04	CCA GTA ATA AAA GGG AGA AAA ATC TAT AAT CA
Neighbor_05	CGG CAC CGC TTC TGG TGA AAT TGT AAA CGT TA
Neighbor_06	GTG AGG CCG AAC GGT ACG CCA GAA AAA GGA AGG GAA GAA A
Neighbor_07	AAC GGG GTC AGT GCC TTT TGA ATA CGC AGA GGC GAA TTA TGT CA
Overhang_01	AACATCGCGATTAGAGCCGTGAGAGATCTACA
Overhang_02	ATATTTTGGAAAAGCCCCAAAAACTAAATGCA
Overhang_03	ATTGCAACAGGAAAAAGTAGCAATGGCGGTCA
Overhang_04	CAGGTCATTGCCTGAGTAGCTGATAAATTAAT
Overhang_05	TAAATCAGATGTCAATCATATGTAGATTCAAA
Overhang_06	GGGTACCGAATGAGTGAGCTAACTCTTCACCG
Overhang_07	TGATATTCAAGCCTCAGAGCATAAGGGCGCGA
Overhang_08	CCAATAGGAGTTTTGTCGTCTTTCTCAACAGT
Overhang_09	ATTTTTCAATAGATAATACATTTGGAACCTAC
Overhang_10	ACTTTTGCGGGAGAAGCCTTTATTTTAGTTTG
Overhang_11	GATGATACGTTTGCCTTTAGCGTCCAATGAAA
Overhang_12	CCTGTTTACCGTACTCAGGAGGTTTACCGTAA
Overhang_13	AACGTAGACCGATTGACCAGTAGCACCATTAC
Overhang_14	TTTTCGAGACCGCACTCATCGAGACGAGGCGT
Overhang_15	AACGCGCCAAAAATAATATCCCATATCTTACC
Overhang_16	CGGTATTCAATAGCAGCCTTTACATGAACACC
Overhang_17	AAATCAAGAACAGCCATATTATTTTTGAGTTA
Overhang_18	GCAACAGCTTGGAACAAGAGTCCAAACGAACT
Overhang_19	CCCAAATCAGAGGACAGATGAACGACTAAAAC
Overhang_20	CCGCCGCGCAAATCAAGTTTTTTGATAAAGTA
Overhang_21	TGTAGCGGAAGCACTAAATCGGAACAGAGGCA
Overhang_22	CCAAATAAGAAACGATTTTTTGTTAGCGCTAA
Overhang_23	TTTTACATCCCCTGCCTATTTCGGGCGCAGTC
Overhang_24	AAATAACAGGAGGCCGATTACTAGGGCGCTGGCAAG
Overhang_25	GAAGTCTGCCAGTTTGAGGGCGCATTAACATGGTCA
Overhang_26	TCCGATATAAAAACCAGCGCAACGTCACAGACTGTA
Overhang_27	ATACGTGTACAGACCAGGCGTTGACAAGATAAGTCC
Overhang_28	AGAATGGGTCGAGGTGCCGTATCACGCTGATCAGAGC
Overhang_29	ATCAAAGAGAATCGATGAACGGCAAATCACCATCAATA
Overhang_30	AACGCTCAGAGAATCGGGGGAAAG
Overhang_31	TCGCCTGATAAATTGTGTCGAAAT
Overhang_32	TGGTTCCGGCAGGCGACACACAAC
Overhang_33	CAGAAGGACATTTTGCTGAACCTC
Overhang_34	AGAATCAAAGGAGTGTACTGGTAA
Overhang_35	AGTAAATGCCCTCATTGAACCGCC
Overhang_36	AAAATTCATGAATTTATCAAAATC
Overhang_37	CTGCGAACCCATATAACAGTTGAT
Overhang_38	TGCTCCATGTAAATTGGGCTTGAG

Overhang_39	AACATAAACAAATCAGATCATTAC
Overhang_40	CAGAAGCAGCCTCAGGAAGATCGCACTCCAGC
Overhang_41	AGGATAAAATATATTTAGGAAGATTGTATAAG
Overhang_42	GAGAGACTTATTTACATTGGCAGATTCACCAG
Overhang_43	TACAGAGGCTTTGAGGGGTCTTTACCCTGACT
Overhang_44	TTTTGCTAAAAGACAGCATCGGAACGAGGGTA

Table A 2: Scrambled overhang sequences

Sequence Name	**Sequence**
Scrambled_01	AACGCTCAGAGAATCGGGGGAAAGTTTTTCCTTAAGAAGTGTATCACCAAT
Scrambled_02	TCGCCTGATAAATTGTGTCGAAATTTTTTGCTCATAATTCTACAAGAGTAC
Scrambled_03	TGGTTCCGGCAGGCGACACACAACTTTTTTAATACACAATGCTACGATCTG
Scrambled_04	CAGAAGGACATTTTGCTGAACCTCTTTTTTGTAAGATACTAGCCCATCTAA
Scrambled_05	AGAATCAAAGGAGTGTACTGGTAATTTTTCGATATGTCAATGCTACCTAAA
Scrambled_06	AGTAAATGCCCTCATTGAACCGCCTTTTTGTACCAATATATAGCAGCTTCA
Scrambled_07	AAAATTCATGAATTTATCAAAATCTTTTTTCAAAGTTAACCGATAGTCTAC
Scrambled_08	CTGCGAACCCATATAACAGTTGATTTTTTATTTGGAACTTACAACTGCACA
Scrambled_09	TGCTCCATGTAAATTGGGCTTGAGTTTTTTGCGATGAATAAACTCTTCACA
Scrambled_10	AACATAAACAAATCAGATCATTACTTTTTAAGATCGCTTACCTATAGATCA
Scrambled_11	CAGAAGCAGCCTCAGGAAGATCGCACTCCAGCTTTTTCTAACTCTAGAATCTCTGAGAA
Scrambled_12	AGGATAAAATATATTTAGGAAGATTGTATAAGTTTTTAAAGTTACCTCCGAAATGTCTA
Scrambled_13	GAGAGACTTATTTACATTGGCAGATTCACCAGTTTTTATCACTTAGATCTACAACGGAT
Scrambled_14	TACAGAGGCTTTGAGGGGTCTTTACCCTGACTTTTTTCGTAAAACCAATTTAGACGTTC
Scrambled_15	TTTTGCTAAAAGACAGCATCGGAACGAGGGTATTTTTACATAGCACACTATCATGGTAT
Scrambled_16	AACATCGCGATTAGAGCCGTGAGAGATCTACATTTTTATCAGACTACCATAGACTTTAG
Scrambled_17	ATATTTTGGAAAAGCCCCAAAAACTAAATGCATTTTTCATTACTTGAGCCGAATATACA
Scrambled_18	ATTGCAACAGGAAAAAGTAGCAATGGCGGTCATTTTTAAATTGTCCAAGCTTGAACTAC
Scrambled_19	CAGGTCATTGCCTGAGTAGCTGATAAATTAATTTTTTTCTTACACCCAGAATGATAATG
Scrambled_20	TAAATCAGATGTCAATCATATGTAGATTCAAATTTTTTCAGTAGACCTCTAAAATGCAT
Scrambled_21	GGGTACCGAATGAGTGAGCTAACTCTTCACCGTTTTTGAAAAGCTGCATTCTCTCATAA
Scrambled_22	TGATATTCAAGCCTCAGAGCATAAGGGCGCGATTTTTCTGACTGTAAACTATAGACCTA
Scrambled_23	CCAATAGGAGTTTTGTCGTCTTTCTCAACAGTTTTTTCGTTCACAGAAGTCTTATACAA
Scrambled_24	ATTTTTCAATAGATAATACATTTGGAACCTACTTTTTTTATACACGTCCTTAAGACAAG
Scrambled_25	ACTTTTGCGGGAGAAGCCTTTATTTTAGTTTGTTTTTTTCAATCACGTACAGCAGTATA
Scrambled_26	GATGATACGTTTGCCTTTAGCGTCCAATGAAATTTTTTAAGACATCTAACTTGACTCGA
Scrambled_27	CCTGTTTACCGTACTCAGGAGGTTTACCGTAATTTTTTTACGATCTACTGTCGACAAAA
Scrambled_28	AACGTAGACCGATTGACCAGTAGCACCATTACTTTTTACGAACTCGGATAACTTTCAAT
Scrambled_29	TTTTCGAGACCGCACTCATCGAGACGAGGCGTTTTTTTCCGAACTAAGTATGAACTTAC
Scrambled_30	AACGCGCCAAAAATAATATCCCATATCTTACCTTTTTCCACTCGATTGTAAGCAAAATT
Scrambled_31	CGGTATTCAATAGCAGCCTTTACATGAACACCTTTTTGAAGTACCTTTAAGATACTACC

Scrambled_32	AAATCAAGAACAGCCATATTATTTTTGAGTTATTTTTCCGAATTTCTACAAACGATTGA
Scrambled_33	GCAACAGCTTGGAACAAGAGTCCAAACGAACTTTTTTATTGACTAACCGATAGACCATT
Scrambled_34	CCCAAATCAGAGGACAGATGAACGACTAAAACTTTTTGACCTAACCAGTTAAGTTACTA
Scrambled_35	CCGCCGCGCAAATCAAGTTTTTTGATAAAGTATTTTTGCCATACTAGAGTACTTTCAAA
Scrambled_36	TGTAGCGGAAGCACTAAATCGGAACAGAGGCATTTTTATCTGAAGATATACACCGCTTA
Scrambled_37	CCAAATAAGAAACGATTTTTTGTTAGCGCTAATTTTTTCGCACGAAAACTAGTATTCTA
Scrambled_38	TTTTACATCCCCTGCCTATTTCGGGCGCAGTCTTTTTACCAAGTAAACTCGATTACTTG
Scrambled_39	AAATAACAGGAGGCCGATTACTAGGGCGCTGGCAAGTTTCAAGCATCGAATGATTCTACA
Scrambled_40	GAAGTCTGCCAGTTTGAGGGCGCATTAACATGGTCATTTCCAGAACCAGTGTTCAATAAT
Scrambled_41	TCCGATATAAAAACCAGCGCAACGTCACAGACTGTATTAACCAGCATTACTTAAGGTATC
Scrambled_42	ATACGTGTACAGACCAGGCGTTGACAAGATAAGTCCTTAGCAGATCTGATACCATAATTC
Scrambled_43	AGAATGGGTCGAGGTGCCGTATCACGCTGATCAGAGCTCTCATGAACACAATGGCTTATA
Scrambled_44	ATCAAAGAGAATCGATGAACGGCAAATCACCATCAATATTACAAGACAACTTTACCTGGA

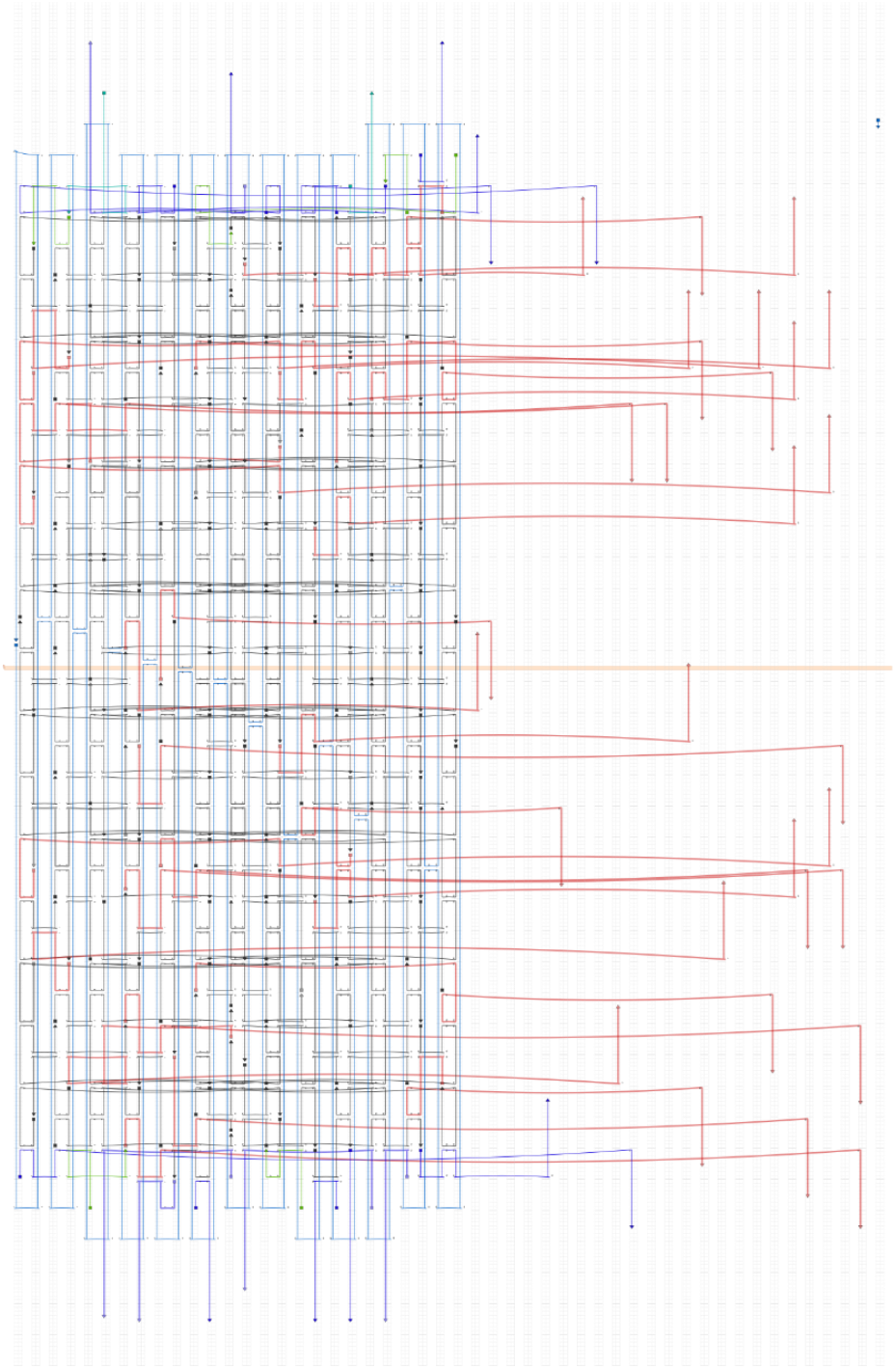

Figure A 1: Staple routing of the Branch.

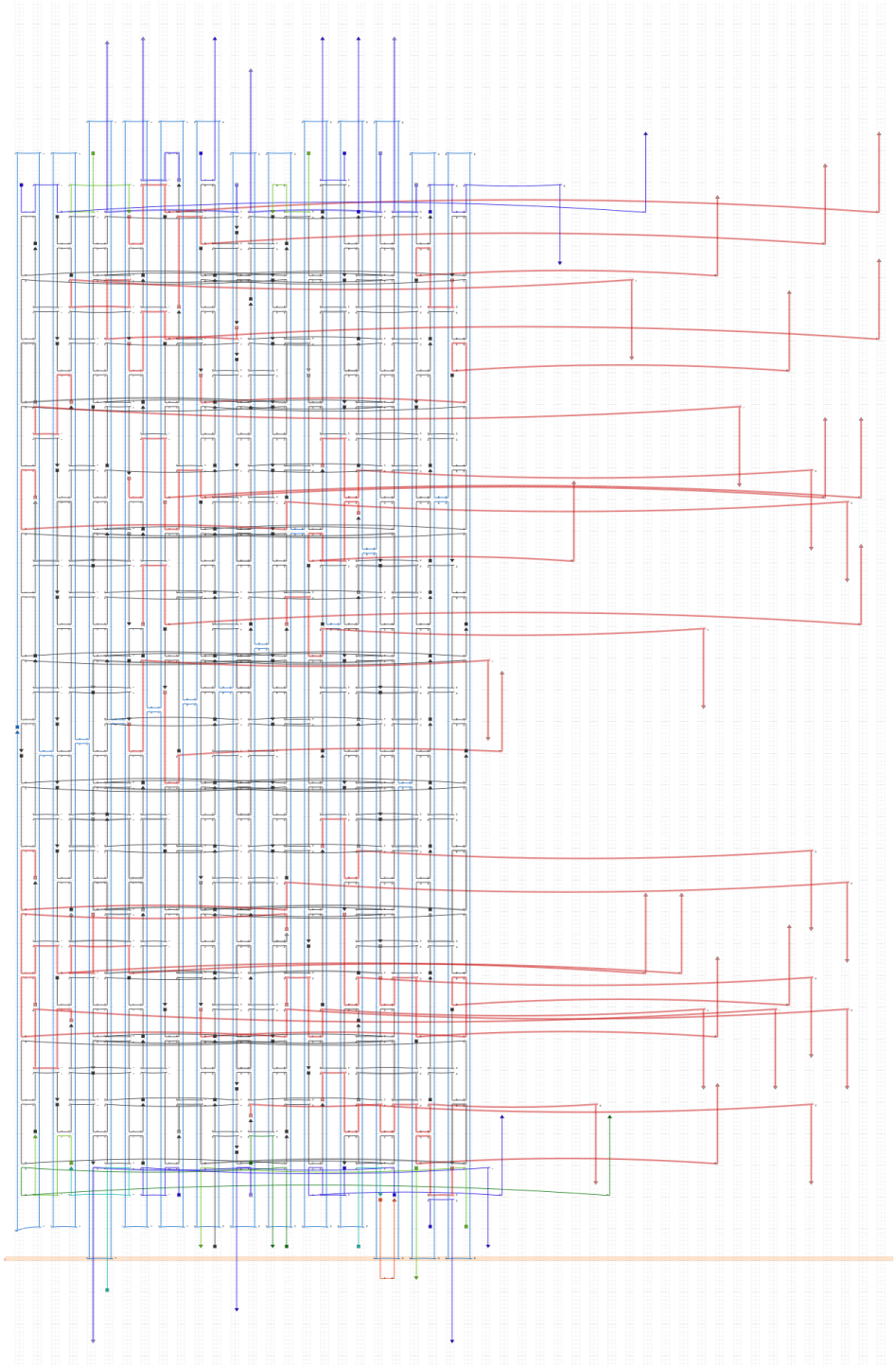

Figure A 2: Staple routing of the modified Branch for project 2.

Appendix B

```
% Halley megafold paper
% plotting and quantifying gel images rapid fold and kinetics

clc, clear all, close all

gel_gray = imread('20150806 Br R7 overhang incorporation test-532nm.tif');
gel_double = im2double(gel_gray);

rect0 = [135.5100    0.5100  488.9800  103.9800];
h1 = figure(1);
imshow(gel_double)
xlabel('crop image to structure lanes','FontSize',20,'Fontweight','bold')
gel_crop_rect = imrect(gca, rect0);
pause
rect = gel_crop_rect.getPosition;
xpos = rect(1);
ypos = rect(2);
box_width = rect(3);
box_height = rect(4);

gel_crop_im = gel_double(ypos:(ypos+box_height),xpos:(xpos+box_width));

%% background subtraction
figure(2)
contour(gel_crop_im,'Fill','on')
set(gcf,'Position',[25 25 400 400])
xlabel('select 30 points for background','FontSize',20,'Fontweight','bold')

fit_length = 20;
x_fit = zeros(fit_length,1);
y_fit = zeros(fit_length,1);
z_fit = zeros(fit_length,1);

% uncomment to pick points from image
for i=1:30
    [x_pt, y_pt] = ginput(1);
    x_fit(i) = round(x_pt);
    y_fit(i) = round(y_pt);
    z_fit(i) = gel_crop_im(y_fit(i),x_fit(i));
    hold on
    plot(x_fit(i),y_fit(i),'kx','Linewidth',2,'MarkerSize',10)
end

% coodinates for horse screen 2
% x_fit = [3 74 158 255 361 425 487 8 128 246 365 485 65 126 251 372 480 5 125 249 372
482 485 249 367 65 10 7 320 427]';
% y_fit = [97 97 100 101 102 101 101 80 79 80 80 82 58 57 57 59 61 38 40 40 41 40 24
25 25 25 22 11 15 15]';
% hold on
% plot(x_fit,y_fit,'kx','Linewidth',2,'MarkerSize',10)
% z_fit = [0.1176 0.1373 0.1569 0.1686 0.1529 0.1255 0.1020 0.1216 0.1529 0.1647
0.1490 0.1137 0.1333 0.1451 0.1333 0.1216 0.1451 0.1059 0.1294 0.1176 0.1020 0.0980
0.0824 0.1098 0.0941 0.1176 0.1098 0.0980 0.0980 0.0863]';

poly3 = polyfitn([x_fit y_fit],z_fit,3);
[r, c] = size(gel_crop_im);

[x_grid, y_grid] = meshgrid(1:c,1:r);
x_bg1 = reshape(x_grid,r*c,1);
```

```
y_bg1 = reshape(y_grid,r*c,1);
z_bg1 = polyvaln(poly3,[x_bg1 y_bg1]);

x_bg = reshape(x_bg1,r,c);
y_bg = reshape(y_bg1,r,c);
z_bg = reshape(z_bg1,r,c);

figure(3)
surf(gel_crop_im,'EdgeColor','none','FaceLighting','gouraud','EdgeLighting','gouraud')
view([0 1 0.5])
set(gcf,'Position',[450 25 400 400])
x_lim = get(gca,'Xlim');
y_lim = get(gca,'Ylim');
z_lim = get(gca,'Zlim');
xlabel('Pre background subtration','FontSize',20,'Fontweight','bold')
set(gca,'Clim',[min(min(gel_crop_im)) max(max(gel_crop_im))])

figure(4)
surf(z_bg,'EdgeColor','none','FaceLighting','gouraud','EdgeLighting','gouraud')
view([0 1 0.5])
set(gcf,'Position',[875 25 400 400])
set(gca,'Xlim',x_lim,'Ylim',y_lim,'Zlim',z_lim)
xlabel('Fitted background','FontSize',20,'Fontweight','bold')
set(gca,'Clim',[min(min(gel_crop_im)) max(max(gel_crop_im))])

gel_norm = gel_crop_im - z_bg;

figure(5)
surf(gel_norm,'EdgeColor','none','FaceLighting','gouraud','EdgeLighting','gouraud')
view([0 1 0.5])
set(gcf,'Position',[1300 25 400 400])
xlabel('Post background subtration','FontSize',20,'Fontweight','bold')

%% Thesholding and rotating image

% threshold image
z_norm = reshape(gel_norm,1,r*c);
im_avg = mean(z_norm);
im_std = std(z_norm);
im_thrsh = zeros(size(gel_norm));
im_thrsh(gel_norm>(im_avg+1*im_std)) = 1;

figure(7)
imshow(im_thrsh)
% imwrite(im_thrsh,'im_thresh.tif','tif');
im_thrsh2 = medfilt2(im_thrsh,[4 4]);
xlabel('Thresholded image','FontSize',20,'Fontweight','bold')
figure(9)
imshow(im_thrsh2)
xlabel('Median Filtered threshold','FontSize',20,'Fontweight','bold')

% rotating image if necessary

theta_deg = -5:0.2:5;
col_num_zero = 1000*ones(size(theta_deg));
max_col_sum = zeros(size(theta_deg));

% uncomment below to do automatic rotation
% for n=1:length(theta_deg)
%     gel_rot_thrsh = imrotate(im_thrsh2,theta_deg(n),'bicubic');
% %     figure(8)
% %     imshow(gel_rot_thrsh)
%     col_sum = sum(gel_rot_thrsh,1);
```

```
%     max_col_sum(n) = max(col_sum);
% %     figure(9)
% %     plot(1:length(col_sum),col_sum)
%     col_nonzeros = nonzeros(col_sum);
%     col_num_zero(n) = length(col_sum)-length(col_nonzeros);
% %     keyboard
% end

% [max_col, max_n] = max(col_num_zero);
% theta_rot = theta_deg(max_n);

% uncomment below to do manual rotation
% manually rotate
figure(8)
imshow(gel_norm,[0 max(max(gel_norm))])
colormap jet
xlabel('Set line for rotation','FontSize',20,'Fontweight','bold')
rot_line = imline(gca);
pause
rot_pos = rot_line.getPosition;
rot_x1 = rot_pos(1,1);
rot_x2 = rot_pos(2,1);
rot_y1 = rot_pos(1,2);
rot_y2 = rot_pos(2,2);
theta_rot = 180/pi*atan((rot_y2-rot_y1)/(rot_x2-rot_x1));



gel_rot_thrsh = imrotate(im_thrsh2,theta_rot,'bicubic');
[r1 c1] = size(gel_rot_thrsh);


%% increasing display resolution by interpolation
gel_rot = imrotate(gel_norm,theta_rot,'bicubic');
gel_norm2 = imresize(gel_rot,4,'bicubic');

figure(6)
surf(gel_norm2,'EdgeColor','none','FaceLighting','gouraud','EdgeLighting','gouraud')
grid off
% view([1 0 4])
view([0 1 20])
set(gcf,'Position',[1725 25 400 400])
set(gca,'XDir','reverse')
xlabel('Rotated image','FontSize',20,'Fontweight','bold')

%% Detect lane edges
N = 7; % number of bands
band_left = zeros(1,N);
band_right = zeros(1,N);

% k_left=1;
% k_right=1;
% k=1; % if k=1 looking for left edge (if k=2 looking for right edge)
%
% for n=1:c1
%     col_sum_n = sum(gel_rot_thrsh(:,n));
%     if k==1 % looking for left edge
%         if col_sum_n ~= 0
%             band_left(k_left)=n-1;
%             k_left=k_left+1;
%             k=2;
%         end
%     else
```

```
%         if col_sum_n==0
%             band_right(k_right)=n;
%             k_right=k_right+1;
%             k=1;
%         end
%     end
% end

h10 = figure(10);
h10_axes = axes;
imshow(gel_rot), hold on
colormap jet
set(gcf,'Position',[750 600 1000 300])



%% Removing negative intensity from wells

% Set a rectangle to define new background. Wells will be removed by
% eliminating any pixels with intensity less than then minimum of the
% background.
xlabel('Set rectangle for background','FontSize',20,'Fontweight','bold')
gel_back_rect = imrect(gca); % selecting rectangle of background on background
subtracted image
back_rect = gel_back_rect.getPosition;
xpos_bg = back_rect(1);
ypos_bg = back_rect(2);
box_width_bg = back_rect(3);
box_height_bg = back_rect(4);

im_back_rect =
gel_rot(ypos_bg:(ypos_bg+box_height_bg),xpos_bg:(xpos_bg+box_width_bg));

% Now plot lane edges after selectingbackground
% for i=1:N
%     plot(band_left(i)*ones(1,r1),1:r1,'r','linewidth',2)
%     plot(band_right(i)*ones(1,r1),1:r1,'b','linewidth',2)
% end

v_line = imline(h10_axes);
v_pos = v_line.getPosition;

for i=1:N
    xlabel('Set left edge','FontSize',20,'Fontweight','bold')
    v_line1 = imline(h10_axes, v_pos);
    pause
    band_left_pos = v_line1.getPosition;
    band_left(i) = band_left_pos(1);
%     v_pos(1,:) = v_pos(1,:)+50;
%     v_pos(2,:) = v_pos(2,:)+50;

    xlabel('Set right edge','FontSize',20,'Fontweight','bold')
    v_line2 = imline(h10_axes, v_pos);
    pause
    band_right_pos = v_line2.getPosition;
    band_right(i) = band_right_pos(1);
%     v_pos(1,:) = v_pos(1,:)+20;
%     v_pos(2,:) = v_pos(2,:)+20;
end

min_bg = min(min(im_back_rect));
gel_rot2 = gel_rot;
gel_rot2(gel_rot2<min_bg)=0;
```

```
figure(11)
surf(gel_rot2,'EdgeColor','none','FaceLighting','gouraud','EdgeLighting','gouraud')
grid off
% view([1 0 4])
view([0 1 20])
set(gcf,'Position',[2225 25 400 400])
set(gca,'XDir','reverse')
xlabel('Neg. intensity of wells removed','FontSize',20,'Fontweight','bold')

%% Calculating amount in folded structure band

% summing total intensity for each lane
% tot_band_int = zeros(1,N);
% for i=1:N
%     tot_band_int(i) = sum(sum(gel_rot(:,band_left(i):band_right(i))));
% end

% Defining section for folded structure band
figure(10)
xlabel('Set top of folded band region','FontSize',20,'Fontweight','bold')
h_line1 = imline(h10_axes);
pause
pos_line1 = h_line1.getPosition;
c1 = polyfit(pos_line1(:,1)',pos_line1(:,2)',1);

pos_line2 = pos_line1;
pos_line2(:,2) = pos_line2(:,2)+14;
xlabel('Set bottom of folded band region','FontSize',20,'Fontweight','bold')
h_line2 = imline(h10_axes,pos_line2);
pause
pos_line2 = h_line2.getPosition;
c2 = polyfit(pos_line2(:,1)',pos_line2(:,2)',1);
%%
% summing intensity in folded structure band for each lane
band_sum = ones(1,N);
for i=1:N
    band_im = gel_rot(:,(band_left(i)):(band_right(i)));
    band_width = band_right(i)-band_left(i);
    figure(12),clf
    subplot(1,2,1)
    imshow(band_im)
    set(gcf,'Position',[1800 600 100 300])
    x1_top = 1;
    y1_top = c1(1)*band_left(i)+c1(2);
    x2_top = band_right(i)-band_left(i)+1;
    y2_top = c1(1)*band_right(i)+c1(2);
    x1_bot = 1;
    y1_bot = c2(1)*band_left(i)+c2(2);
    x2_bot = band_right(i)-band_left(i)+1;
    y2_bot = c2(1)*band_right(i)+c2(2);
%     hold on
%     plot([x1_top x2_top x2_bot x1_bot x1_top],[y1_top y2_top y2_bot y1_bot
y1_top],'r--')
%     pause
    poly_band = impoly(gca,[x1_top, y1_top; x2_top, y2_top; x2_bot, y2_bot; x1_bot,
y1_bot],'Closed',1);
    band_mask = poly_band.createMask;
    band_sum_im = band_im.*band_mask;
    band_sum(i) = sum(sum(band_sum_im));
end

%%
```

```
lane = 1:N;
Temp = [0 .25 .5 1 2.5 5 10];
% Temp = fliplr([40 41.7 44.4 47.8 52.5 56 58.4 60]);
% Temp = fliplr([48 48.4 48.9 49.6 50.5 51.2 51.7 52]); % LPP fine screen
% Temp = fliplr([50.0 50.4 50.9 51.6 52.5  53.2 53.7 54.0]); % 18hb fine screen
% Temp = fliplr([58 58.4 58.9 59.6 60.5 61.2 61.7 62]); % Horse fine screen
% band_sum_norm = band_sum;

figure(13),hold on, box on
set(gcf,'Color',[1 1 1])
set(gca,'FontSize',30,'Xlim',[min(Temp) max(Temp)],'Ylim',[0 320])
plot(Temp,band_sum,'k','linewidth',2)
xlabel('Excess cold miR','FontSize',30)
ylabel('Intensity (a.u.)','FontSize',30)



% figure(14),hold on, box on
% set(gcf,'Color',[1 1 1])
% set(gca,'FontSize',30,'Xlim',[min(Temp) max(Temp)],'Xdir','reverse','Ylim',[-0.05
1])
% plot(Temp,band_sum_norm,'k','linewidth',2)
% xlabel('Annealing Temperature (^oC)','FontSize',30)
% ylabel('Lane Normalized Intensity','FontSize',30)

filename = '20150806 Br R7 overhang incorporation test-532nm.txt';
fileID = fopen(filename,'w+');
for i=1:N
% fprintf(fileID,'%4.2f\t%4.2f\t%4.3f\n',Time(i),band_sum(i),band_sum_norm(i));
fprintf(fileID,'%4.2f\t%4.2f\n',Temp(i),band_sum(i));
end
fclose(fileID);
```

Appendix C

```
% This program takes a specific input sequence entered as 'seq' and tests
% it agains a list of staples and/or ssDNA ('stapletext') to find the maximum number
of
% total base pairs and the maximum number of sequential base pairs.  These
% can then be checked in IDT to test the stability.

clear all, clc, close all

tic

seq='AGGAAGCCCUGGAGGGGCUGGAG';

% good sequence for FRET box 'CGGGAAGACACTAGG'

NN=length(seq); % Length of sequence
showall=0;

    input_seq=seq;

    % ComplimentaryStapleCheck.txt should list staple sequences in rows
    stapletext = importdata('ciRS-7 sequence.txt','%s');

    max_comp_tot=zeros(1,length(stapletext));
    max_seq_comp=zeros(1,length(stapletext));
    for QQ = 1:length(stapletext) % check each staple in file

        test_staple=stapletext{QQ};

        % Compute Reverse Compliment
        staple_comp=test_staple;
        staple_comp(test_staple=='A')='U';
        staple_comp(test_staple=='U')='A';
        staple_comp(test_staple=='G')='C';
        staple_comp(test_staple=='C')='G';
        staple_Rcomp=fliplr(staple_comp);

%         sprintf('\n%s\n%s\n',rand_seq,staple_Rcomp)

        % Keep rand_seq fixed and run staple reverse complement across the
        % rand_seq to check for complementary base pairs

        l_stap=length(test_staple);
        l_rand=length(input_seq);

        comp_tot=zeros(1,(l_stap+l_rand-1));
        seq_comp=zeros(1,(l_stap+l_rand-1));
        for i=1:(l_stap+l_rand-1)
            b_stap1=max([(l_stap-i+1),1]); % first staple base in test range
            b_stap2=min([l_stap,(l_rand+l_stap-i)]); % second staple base in test
range
            b_rand1=max([1,(i-l_stap+1)]); % first rand_seq base in test range
            b_rand2=min([l_rand,i]); % second rand_seq base in test range

            % printing out sequences for visualizing comparison
%             clc
%             stap_text=strcat(sprintf(strcat('''%',num2str(i),'s'''),'
'),staple_Rcomp);
%             rand_text=strcat(sprintf(strcat('''%',num2str(l_stap),'s'''),'
'),rand_seq);
```

```
%             sprintf('\n%s\n%s\n',stap_text,rand_text)

            % comparing sequences
            rand_test=input_seq(b_rand1:b_rand2);
            stap_test=staple_Rcomp(b_stap1:b_stap2);
            comp_tot(i)=sum(rand_test==stap_test);
            sum_seq_comp=zeros(size(rand_test));

            % calculating maximum sequential number of complementary base pairs
            for j=1:length(rand_test)
                if j==1
                    if rand_test(j)==stap_test(j)
                        sum_seq_comp(j)=1;
                    end
                else
                    if rand_test(j)==stap_test(j)
                        sum_seq_comp(j)=sum_seq_comp(j-1)+1;
                    else
                        sum_seq_comp(j)=0;
                    end
                end
            end
            seq_comp(i)=max(sum_seq_comp);
%             pause(0.01)
        end

        % maximum total complementary base pairs for the staple/rand_seq combination
        max_comp_tot(QQ)=max(comp_tot);
        i_mct=find(max_comp_tot(QQ)==comp_tot);
        % maximum sequential complementary base pairs for the staple/rand_seq
combination
        max_seq_comp(QQ)=max(seq_comp);
        i_msc=find(max_seq_comp(QQ)==seq_comp);

        if showall
            fprintf('\n\nStaple: %s\nRandon sequence: %s\n',test_staple,input_seq)

            for i=1:length(i_mct)
                b_mct_stap1=max([(l_stap-i_mct(i)+1),1]);
                b_mct_stap2=min([l_stap,(l_rand+l_stap-i_mct(i))]);
                b_mct_rand1=max([1,(i_mct(i)-l_stap+1)]);
                b_mct_rand2=min([l_rand,i_mct(i)]);
                rand_test=input_seq(b_mct_rand1:b_mct_rand2);
                stap_test=staple_Rcomp(b_mct_stap1:b_mct_stap2);
                n_blanks=max([l_stap,i_mct(i)]);
                base_match=rand_test;
                base_match(stap_test==rand_test)='|';
                base_match(stap_test~=rand_test)=' ';
                stapRC_text=strcat(sprintf(strcat('''%',num2str(i_mct(i)),'s'''),'
'),staple_Rcomp);
                stap_text=strcat(sprintf(strcat('''%',num2str(i_mct(i)),'s'''),'
'),fliplr(test_staple));
                bp_text=strcat(sprintf(strcat('''%',num2str(n_blanks),'s'''),'
'),base_match);
                rand_text=strcat(sprintf(strcat('''%',num2str(l_stap),'s'''),'
'),input_seq);
                fprintf('\nMax total complementary base pairs is:
%s\n\n%s\n%s\n%s\n%s\n',num2str(max_comp_tot(QQ)),stap_text,stapRC_text,bp_text,rand_t
ext)
%                 pause
            end

            for i=1:length(i_msc)
```

```
                b_msc_stap1=max([(l_stap-i_msc(i)+1),1]);
                b_msc_stap2=min([l_stap,(l_rand+l_stap-i_msc(i))]);
                b_msc_rand1=max([1,(i_msc(i)-l_stap+1)]);
                b_msc_rand2=min([l_rand,i_msc(i)]);
                rand_test=input_seq(b_msc_rand1:b_msc_rand2);
                stap_test=staple_Rcomp(b_msc_stap1:b_msc_stap2);
                n_blanks=max([l_stap,i_msc(i)]);
                base_match=rand_test;
                base_match(stap_test==rand_test)='|';
                base_match(stap_test~=rand_test)=' ';
                stapRC_text=strcat(sprintf(strcat('''%',num2str(i_msc(i)),'s'''),'
'),staple_Rcomp);
                stap_text=strcat(sprintf(strcat('''%',num2str(i_mct(i)),'s'''),'
'),fliplr(test_staple));
                bp_text=strcat(sprintf(strcat('''%',num2str(n_blanks),'s'''),'
'),base_match);
                rand_text=strcat(sprintf(strcat('''%',num2str(l_stap),'s'''),'
'),input_seq);
                fprintf('\nMax sequential complementary base pairs is:
%s\n\n%s\n%s\n%s\n%s\n',num2str(max_seq_comp(QQ)),stap_text,stapRC_text,bp_text,rand_t
ext)
%                 pause
            end
        end

%         pause
    end
    clear comp_tot

    max_all_staple_comp_tot=max(max_comp_tot);
    i_mast=find(max_all_staple_comp_tot==max_comp_tot);
    max_all_staple_seq_comp=max(max_seq_comp);
    i_mass=find(max_all_staple_seq_comp==max_seq_comp);

    %% Repeating loop for the staples that yield maximum number of complmentary base
pairs

    fprintf('\nMax total complementary base pairs is: %i\n',max_all_staple_comp_tot)
    for nn=1:length(i_mast)
        test_staple=stapletext{i_mast(nn)};

        % Compute Reverse Compliment
        staple_comp=test_staple;
        staple_comp(test_staple=='A')='U';
        staple_comp(test_staple=='U')='A';
        staple_comp(test_staple=='G')='C';
        staple_comp(test_staple=='C')='G';
        staple_Rcomp=fliplr(staple_comp);

%         sprintf('\n%s\n%s\n',rand_seq,staple_Rcomp)

        % Keep rand_seq fixed and run staple reverse complement across the
        % rand_seq to check for complementary base pairs

        l_stap=length(test_staple);
        l_rand=length(input_seq);

        comp_tot=zeros(1,(l_stap+l_rand-1));
        seq_comp=zeros(1,(l_stap+l_rand-1));
        for i=1:(l_stap+l_rand-1)
            b_stap1=max([(l_stap-i+1),1]); % first staple base in test range
            b_stap2=min([l_stap,(l_rand+l_stap-i)]); % second staple base in test
range
```

```
            b_rand1=max([1,(i-l_stap+1)]); % first rand_seq base in test range
            b_rand2=min([l_rand,i]); % second rand_seq base in test range

            % comparing sequences
            rand_test=input_seq(b_rand1:b_rand2);
            stap_test=staple_Rcomp(b_stap1:b_stap2);
            comp_tot(i)=sum(rand_test==stap_test);

%             pause(0.01)
        end

        % maximum total complementary base pairs for the staple/rand_seq combination
        max_comp_tot1=max(comp_tot);
        i_mct1=find(max_comp_tot1==comp_tot);

        fprintf('\n\nStaple #%i: %s\nTest sequence:
%s\n',i_mast(nn),test_staple,input_seq)

        for i=1:length(i_mct1)
            b_mct_stap1=max([(l_stap-i_mct1(i)+1),1]);
            b_mct_stap2=min([l_stap,(l_rand+l_stap-i_mct1(i))]);
            b_mct_rand1=max([1,(i_mct1(i)-l_stap+1)]);
            b_mct_rand2=min([l_rand,i_mct1(i)]);
            rand_test=input_seq(b_mct_rand1:b_mct_rand2);
            stap_test=staple_Rcomp(b_mct_stap1:b_mct_stap2);
            n_blanks=max([l_stap,i_mct1(i)]);
            base_match=rand_test;
            base_match(stap_test==rand_test)='|';
            base_match(stap_test~=rand_test)=' ';
            stapRC_text=strcat(sprintf(strcat('''%',num2str(i_mct1(i)),'s'''),'
'),staple_Rcomp);
            stap_text=strcat(sprintf(strcat('''%',num2str(i_mct1(i)),'s'''),'
'),fliplr(test_staple));
            bp_text=strcat(sprintf(strcat('''%',num2str(n_blanks),'s'''),'
'),base_match);
            rand_text=strcat(sprintf(strcat('''%',num2str(l_stap),'s'''),'
'),input_seq);
            fprintf('\n%s\n%s\n%s\n%s\n',stap_text,stapRC_text,bp_text,rand_text)
%             pause
        end
    end

    %% Repeating loop for staples that yield maximum number of sequential
complementary bases

    fprintf('\nMax sequential complementary base pairs is:
%i\n',max_all_staple_seq_comp)
    for nn=1:length(i_mass)
        test_staple=stapletext{i_mass(nn)};

        % Compute Reverse Compliment
        staple_comp=test_staple;
        staple_comp(test_staple=='A')='U';
        staple_comp(test_staple=='U')='A';
        staple_comp(test_staple=='G')='C';
        staple_comp(test_staple=='C')='G';
        staple_Rcomp=fliplr(staple_comp);

%         sprintf('\n%s\n%s\n',rand_seq,staple_Rcomp)

        % Keep rand_seq fixed and run staple reverse complement across the
        % rand_seq to check for complementary base pairs
```

```
        l_stap=length(test_staple);
        l_rand=length(input_seq);

        seq_comp=zeros(1,(l_stap+l_rand-1));
        for i=1:(l_stap+l_rand-1)
            b_stap1=max([(l_stap-i+1),1]); % first staple base in test range
            b_stap2=min([l_stap,(l_rand+l_stap-i)]); % second staple base in test
range
            b_rand1=max([1,(i-l_stap+1)]); % first rand_seq base in test range
            b_rand2=min([l_rand,i]); % second rand_seq base in test range

            % comparing sequences
            rand_test=input_seq(b_rand1:b_rand2);
            stap_test=staple_Rcomp(b_stap1:b_stap2);
            sum_seq_comp=zeros(size(rand_test));

            % calculating maximum sequential number of complementary base pairs
            for j=1:length(rand_test)
                if j==1
                    if rand_test(j)==stap_test(j)
                        sum_seq_comp(j)=1;
                    end
                else
                    if rand_test(j)==stap_test(j)
                        sum_seq_comp(j)=sum_seq_comp(j-1)+1;
                    else
                        sum_seq_comp(j)=0;
                    end
                end
            end
            seq_comp(i)=max(sum_seq_comp);
%             pause(0.01)
        end

        % maximum total complementary base pairs for the staple/rand_seq combination
        max_seq_comp1=max(seq_comp);
        i_msc1=find(max_seq_comp1==seq_comp);

        fprintf('\n\nStaple #%i: %s\nRandom sequence:
%s\n',i_mass(nn),test_staple,input_seq)
        for i=1:length(i_msc1)
            b_msc_stap1=max([(l_stap-i_msc1(i)+1),1]);
            b_msc_stap2=min([l_stap,(l_rand+l_stap-i_msc1(i))]);
            b_msc_rand1=max([1,(i_msc1(i)-l_stap+1)]);
            b_msc_rand2=min([l_rand,i_msc1(i)]);
            rand_test=input_seq(b_msc_rand1:b_msc_rand2);
            stap_test=staple_Rcomp(b_msc_stap1:b_msc_stap2);
            n_blanks=max([l_stap,i_msc1(i)]);
            base_match=rand_test;
            base_match(stap_test==rand_test)='|';
            base_match(stap_test~=rand_test)=' ';
            stapRC_text=strcat(sprintf(strcat('''%',num2str(i_msc1(i)),'s'''),'
'),staple_Rcomp);
            stap_text=strcat(sprintf(strcat('''%',num2str(i_msc1(i)),'s'''),'
'),fliplr(test_staple));
            bp_text=strcat(sprintf(strcat('''%',num2str(n_blanks),'s'''),'
'),base_match);
            rand_text=strcat(sprintf(strcat('''%',num2str(l_stap),'s'''),'
'),input_seq);
            fprintf('\n%s\n%s\n%s\n%s\n',stap_text,stapRC_text,bp_text,rand_text)
%             pause
        end
```

```
    end
%     pause(2)

fprintf('\n\n****end****\n\n')
%seq
% overall_max_comp_consec

toc
```